

THE
*Philosophical
Review*

VOLUME 122, NUMBER 3, JULY 2013

Marygrove College Library
8425 West McNichols Road
Detroit, MI 48221

Published for
the Sage School of Philosophy at Cornell University
by Duke University Press

The *Philosophical Review* is edited by the faculty of the Sage School of Philosophy at Cornell University: Karen Bennett, Richard Boyd, Tad Brennan, Andrew Chignell, Matti Eklund, Gail Fine, Carl Ginet, Harold Hodes, T. H. Irwin, Michelle Kosch, Scott MacDonald, Richard W. Miller, Jill North, Derk Pereboom, Sydney Shoemaker, Ted Sider, Nicholas Silins, Will Starr, Nicholas L. Sturgeon, and Erin Taylor. The journal's editorial associate is Louise Silberling.

Editors

2013 Matti Eklund, Derk Pereboom

2012–2013 Nicholas L. Sturgeon, Matti Eklund, Derk Pereboom

The journal is published four times a year for the Sage School of Philosophy at Cornell University by Duke University Press. See the inside back cover for submission guidelines.

World Wide Web

Visit the editorial office at www.arts.cornell.edu/philrev and Duke University Press Journals at www.dukeupress.edu/journals.

Subscriptions

Direct all orders to Duke University Press, Journals Customer Service, 905 W. Main St., Suite 18B, Durham, NC 27701. Annual subscription rates: print-plus-electronic institutions, \$165; print-only institutions, \$155; e-only institutions, \$130; individuals, \$33; students, \$20. For information on subscriptions to the e-Duke Journals Scholarly Collections through HighWire Press, see www.dukeupress.edu/library/eDuke. Print subscriptions: add \$14 postage and applicable HST (including 5% GST) for Canada; add \$18 postage outside the US and Canada. Back volumes (institutions): \$155. Single issues: institutions, \$39; individuals, \$12. For more information, contact Duke University Press Journals at 888-651-0122 (toll-free in the US and Canada) or 919-688-5134; subscriptions@dukeupress.edu.

Permissions

Photocopies for course or research use that are supplied to the end user at no cost may be made without explicit permission or fee. Photocopies that are provided to the end user for a fee may not be made without payment of permission fees to Duke University Press. Address requests for permission to republish copyrighted material to Rights and Permissions Manager, permissions@dukeupress.edu.

Advertisements

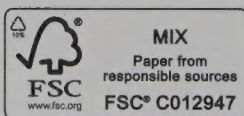
Direct inquiries about advertising to Journals Advertising Coordinator, journals_advertising@dukeupress.edu.

Indexing and Abstracting

For a list of the sources in which *Philosophical Review* is indexed and abstracted, see www.dukeupress.edu/philreview.

ISSN 0031-8108

© 2013 by Cornell University



THE
Philosophical
Review

VOLUME 122, NUMBER 3, JULY 2013

Epistemic Teleology and the Separateness of Propositions

Selim Berker 337

Peer Disagreement, Evidence, and Well-Groundedness

Han van Wietmarschen 395

Truth, the Liar, and Relativism

Kevin Scharp 427

BOOK REVIEWS

Derk Pereboom, *Consciousness and the Prospects of Physicalism*

Christopher S. Hill 511

Casey Perin, *The Demands of Reason: An Essay on Pyrrhonian Scepticism*

Stefan Sienkiewicz 519

Gordon Belot, *Geometric Possibility*

Jill North 522

Epistemic Teleology and the Separateness of Propositions

Selim Berker

Harvard University

1. Introduction

As I see it, the most fundamental question in ethics is “What should I do?”¹ The exact wording of this question can take different forms, depending on one’s normative vocabulary of choice (“What do I have most reason to do?” “What is it rational for me to do?” “What am I justified in doing?”). And the question can take broader or narrower scope, depending on whether it is applied to a specific situation (“Should I

In writing this essay, I have received help from an embarrassingly large number of people. Many thanks to Tyler Doggett, Stephen Grimm, Derek Parfit, Phillip Pettit, T. M. Scanlon, Drew Schroeder, Mark van Roojen, and Dennis Whitcomb for their immensely helpful written comments; to Ruth Chang, David Copp, Tom Dougherty, Catherine Elgin, Richard Feldman, Daniel Greco, Anil Gupta, Elizabeth Harman, Thomas Kelly, Niko Kolodny, John MacFarlane, Amanda Montgomery, Oded Na’aman, Ram Neta, Shantia Rahimian, Joseph Raz, Nishi Shah, Susanna Siegel, Nico Silins, Michael Smith, Daniel Star, Kurt Sylvan, Jonathan Vance, Jonathan Vogel, Gerard Vong, Robbie Williams, and Seth Yalcin for crucial discussion; to audiences at Boston University, Brandeis University, Cornell University, Leeds University, New York University, Princeton’s University Center for Human Values, the University of California, Berkeley, the University of North Carolina at Chapel Hill, the University of Pittsburgh, and the University of Rochester for their incisive questions; to the members of my spring 2011 graduate seminar at Harvard University for their invaluable input; and to three anonymous referees at *Philosophical Review* for their excellent suggestions on ways to improve the essay.

1. More accurately, the fundamental question in ethics is “What, if anything, should I do?” to allow for the possibility that there is no such thing as should-be-done-ness. Since I will not be broaching the subject of normative nihilism in the current essay, I leave this complication to one side.

tell her the truth?") or to one's life as a whole ("What sort of a life should I live?"). One branch of ethics—metaethics—seeks to clarify this fundamental question, in its many different manifestations. What does the question mean? What would constitute an answer to it? How can we go about figuring out that answer? What would the objective status of such an answer be, were we able to find one? The other branch of ethics—normative ethics—offers substantive answers to our fundamental question, in its various guises. Sometimes these answers take the form of grand theories that distill should-be-done-ness down to a pristine set of necessary and sufficient conditions—the familiar conflicting “-isms” of introductory ethics courses (egoism, utilitarianism, Kantianism, contractualism, and the like). Other times these answers are more modest in scale, taking the form of a mid-level principle such as the doctrine of double effect, or even just a particular verdict about a particular scenario.

I think it is useful to view epistemology as having a parallel structure. On this way of conceiving of the discipline, the most fundamental question in epistemology is “What should I believe?” The exact wording of this question can take different forms, depending on one's normative vocabulary of choice (“What do I have most reason to believe?” “What is it rational for me to believe?” “What am I justified in believing?”). And the question can take broader or narrower scope, depending on whether it is applied to a specific proposition (“Should I believe that there is a God?”) or to one's cognitive life as a whole (“How should I go about forming my beliefs?”). One branch of epistemology—metaepistemology, as we might call it—seeks to clarify this fundamental question, in its many different manifestations. The other branch of epistemology—normative epistemology, if you will—offers substantive answers to our fundamental question, in its various guises, resulting in the familiar conflicting “-isms” from introductory epistemology courses (Cartesian foundationalism, coherentism, reliabilism, and the like).

Now this way of viewing epistemology is controversial. For instance, many contemporary epistemologists frame their fundamental question in the third person: “What is *he* or *she* justified in believing?” not “What am *I* justified in believing?” Recently, a growing number of epistemologists have emphasized the social nature of epistemic norms: they ask, in effect, “What do *we* have most reason to believe?” And, most controversially, my way of characterizing the field of epistemology seems to blatantly disregard its etymological origins as “the study (-ology) of knowledge (*epistēmē*)”: for many epistemologists, the fundamental question in epistemology is “What makes a belief count as *knowledge*?” I think

each of these alternate epistemological questions is important in its own right, but I regard them as auxiliary questions that can be addressed only after significant headway has been made on the more fundamental question “What should I believe?”—although I realize that, in saying this, I am making a bold claim that I can’t hope to defend here.

One virtue of viewing epistemology in the manner I have been proposing is that it brings to the fore the *normative* character of the discipline. Moreover, I think that exploring the analogy between ethics and epistemology can help us make progress in both fields. But this thought is compatible with recognizing that, in the end, the tempting parallelism between ethics and epistemology may turn out to be partially illusory. The surface similarity of the questions “What should I do?” and “What should I believe?” might mask a deeper divide between what is really being asked by these questions and what a proper answer to them involves; that is, our metaethics might end up looking very different from our metaepistemology, and our normative ethics might end up looking very different from our normative epistemology. Indeed, I think this result is almost inevitable, at least to a certain degree. After all, belief is not action, and action is not belief. So at some point the analogy between ethics and epistemology must break down. However, the ethics–epistemology analogy is not alone in this regard. *All* analogies break down at some point; that’s what makes them analogies and not identities. So even if we must be careful not to push it too far, the ethics–epistemology analogy can, I think, do real work for us.

In what follows, I want to use this analogy between normative evaluations of actions and normative evaluations of beliefs to tease out, and then argue against, a certain strain of thought that seems to have become an article of faith in much recent epistemological theorizing. According to this strain of thought, what distinguishes epistemic norms from other sorts of norms (prudential, moral, and so on) is that epistemic norms are guided by a distinctive set of “epistemic” or “cognitive” or “intellectual” goals. Most often these epistemic goals are taken to be the twin goals of acquiring true beliefs and avoiding false ones. Sometimes, though, the list of epistemic goals is broadened to include other items, such as the acquisition of knowledge, or coherent belief-systems, or understanding, or wisdom. But regardless of what the list of epistemic goals looks like, the guiding idea behind this strain of thought is that all other normative notions in epistemology are ultimately explicable in terms of how well the objects of assessment conduce toward, promote, or otherwise subserve these epistemic goals.

In a way, this picture is a very natural one, which perhaps explains why it is so widely endorsed (and almost always without being explicitly argued for). However, by thinking about the analogy between ethics and epistemology, we can at once see the degree to which this picture has some quite substantial normative assumptions built into it. According to this picture, there are certain epistemic *ends* or *goals* that it is epistemically *good* for us to promote, and the question of what we *should* believe is determined by how well our believing conduces toward the fulfilling of those goals, or the furthering of those ends. But this, of course, is the analogue in normative epistemology of *consequentialist* or *teleological* approaches to normative ethics: it makes the good prior to the right with regard to the epistemic evaluation of beliefs, just as consequentialist/teleological ethical systems are often characterized as making the good prior to the right with regard to the ethical evaluation of actions. Maybe, in the end, such an approach is justified. But it is hardly an innocent starting point for epistemological theorizing that can just be assumed without argument.

Indeed, my aim in this essay is to convince you of something stronger. I want to convince you that this consequentialist/teleological approach to normative epistemology is positively misguided—that is, I want to convince you that, when providing a substantive answer to the question “What should I believe?” or “What am I justified in believing?” we should not proceed by first identifying certain epistemic goods and then constructing a theory of epistemic justification in terms of what conduces toward or promotes those epistemic goods. My basic argument will be relatively simple. Consequentialist/teleological theories in ethics have a certain structure: in particular, they almost always countenance trade-offs between various goods in the determination of which action should be done. So if the consequentialist/teleological approach to normative epistemology were the correct one, we’d expect that the correct theory of what we should believe would also countenance trade-offs between goods—in this case, trade-offs between epistemic goods. However, no one—not even those epistemologists who most explicitly embrace the consequentialist/teleological framework—is willing to countenance all such trade-offs in the epistemic case. So, I will conclude, this entire approach to normative epistemology is misguided: its advocates don’t realize what their approach really commits them to, and if they did realize it, they would abandon the approach rather than incorporate the commitment. Fully defending this conclusion, though, will take a bit of care since modern consequentialists have developed a variety of techniques

for including bans on at least certain types of trade-offs within a broadly consequentialist ethical framework. Thus I will need to argue that none of these techniques can be employed in the epistemic case—or that if they can be employed, then they don't succeed in prohibiting the sorts of epistemic trade-offs that, I will argue, nearly everyone deems to be unacceptable.

First, though, we should be a bit more precise about what, exactly, a consequentialist or teleological approach involves, whether in the case of ethics or epistemology.

2. Ethical Teleology and Epistemic Teleology

Modern introductory courses in normative ethics tend to have a familiar narrative: the titanic clash between deontological and consequentialist moral theories. On the consequentialist side, one studies act-utilitarianism, nonutilitarian varieties of act-consequentialism, and (if there's time) rule-consequentialism. On the deontological side, one studies various formulations of Kant's categorical imperative, Ross's theory of *prima facie* duties, and (if there's time) contractualism. The familiar refrain is that consequentialist theories define the right in terms of the good, whereas deontological theories reverse that order of explanation.

This division of all ethical theories into two general categories depending on their stand on the explanatory priority of goodness/betterness/value versus rightness/obligation/duty appears to have been a staple of most twentieth-century ethics courses in the analytic tradition if introductory textbooks and anthologies are any guide. The slightly older term for much the same distinction was "teleological" (from the Greek *telos* for 'end') versus "deontological" (from the Greek *deon* for 'duty') ethical theories. This terminology was introduced as a way of partitioning all ethical theories into two groups by C. D. Broad in 1930,² and most

2. "I would first divide ethical theories into two classes, which I will call respectively *deontological* and *teleological*."

Deontological theories hold that there are ethical propositions of the form: 'Such and such a kind of action would always be right (or wrong) in such and such circumstances, no matter what its consequences might be'.... Teleological theories hold that the rightness or wrongness of an action is always determined by its tendency to produce certain consequences which are intrinsically good or bad." Broad 1930, 206–7. Also responsible for popularizing this use of 'teleological' and 'deontological' was J. H. Muirhead (1932, 6), who in his book *Rule and End in Morals* used these words to characterize two opposing "points of view" in ethics. Note that 'teleological' and 'deontological' had somewhat different meanings in ethics before Broad's and Muirhead's use of them

ethics textbooks and anthologies from the 1940s, 1950s, and 1960s that are written in English follow Broad's taxonomy.³ However, the basic distinction appears to predate Broad. For example, in their textbook *Ethics* from 1908, John Dewey and James H. Tufts (1908, 224–25) call moral theories that take value or goodness as their fundamental idea “teleological” and moral theories that take duty or rightness as their fundamental idea “jural” (from the Latin *jus* for ‘law’). Moreover, the manner in which Dewey and Tufts write suggests that this division was not their own creation, but rather one that was common at the time their book was published.⁴

In her essay “Modern Moral Philosophy” from 1958, Elizabeth Anscombe (1958, 12) introduced the term ‘consequentialism’,⁵ and gradually during the latter half of the twentieth century, ‘consequentialist’ came to replace ‘teleological’ as the chief designator of the sort of ethical theory that is said to contrast with deontological ones. But despite the current ubiquity of the term ‘consequentialism’, in this essay I will primarily be using the pre-Anscombian term ‘teleology’ to refer to the type of ethical and epistemic theories under consideration. My main reason for doing so is that the phrase ‘epistemic consequentialism’ is liable to cause confusion in a way that the phrase ‘epistemic teleology’ does not. ‘Epistemic consequentialism’ suggests a theory according to which whether a belief is justified or unjustified depends on the *nonepistemic* goodness or badness of the consequences of believing it; however, that’s not the sort of theory defended by the people who are the target of my discussion here. ‘Epistemic teleology’, on the other hand, more aptly captures the sort of position in normative epistemology that I aim to criticize: namely, an epistemic theory that is *structurally parallel to* teleological/consequentialist theories in normative ethics, not one that is *an instance of* such an ethical theory.

became standard; in fact, ‘deontology’ began its lexical life as Jeremy Bentham’s name for what I will call a teleologist’s deontic theory. For discussion, see Loudon 1996.

3. See, for example, Tsanoff 1947, 38–41; Garvin 1953, 193; Brandt 1961, 131–32; Zink 1962, 154; Frankena 1963, 13–14; and Banner 1968, 167. Tsanoff and Garvin use ‘formalism’ instead of ‘deontology’ as the contrast term for ‘teleological ethics’, whereas Brandt uses both ‘formalism’ and ‘deontology’, which for him are synonyms.

4. In particular, Dewey and Tufts (1908, 224) say that moral theories of the first sort are “frequently” called teleological.

5. Though, as Derek Parfit reminds me, with a different meaning from the one now used. Thus ‘consequentialism’, ‘teleology’, and ‘deontology’ have all shifted in meaning over the years; see n. 2.

Providing a precise characterization of the teleological/deontological distinction is difficult to do. Nearly everyone agrees that hedonistic act-utilitarianism is a canonical example of a teleological ethical theory and that Kantianism is a canonical example of a deontological ethical theory, but this is where the agreement ends. Often authors build into their definition of teleological or consequentialist theories a commitment to the maximization of value,⁶ but this is a mistake, for it neglects nonmaximizing versions of consequentialism. Even the familiar dictum that teleological/consequentialist theories make the good prior to the right and deontological theories make the right prior to the good is problematic. On the teleological/consequentialist side of things, this dictum rules out an increasingly popular form of consequentialism known as “scalar consequentialism,” which holds that there is no such thing as obligation or rightness, just betterness.⁷ On the deontological side of things, the dictum rules out theories according to which sometimes an action’s rightness or wrongness is entirely determined by the goodness of its consequences, but other times its rightness or wrongness is determined by other factors (such as the fact that it violates an absolute side-constraint). Moreover, the familiar dictum also neglects a recent strain of Kant interpretation that takes at face value Kant’s claims about the unconditional goodness of a good will in the opening lines of the *Groundwork* and sees Kantian ethics as founded on a distinctive theory of value.⁸

This last observation points the way toward a more satisfactory definition of teleology. What is distinctive about the teleological perspective is not just its taking value to be fundamental but moreover its attitude toward the nature of value and how we should respond to it. According to the teleologist, the proper response to value is to bring it about, and the proper response to disvalue is to stop it from being brought about: in short, for the teleologist all value is “to be promoted,” and all disvalue is “to be prevented” (Pettit 1991, 230–33; Korsgaard 1993, 24; Pettit 1997, 125–30; Scanlon 1998, 79–80). This leads to two constraints on any given teleological theory. First, the fundamental bearers of value, whatever they

6. See, for example, Frankena 1963, 13; Rawls 1971, 24; Scheffler 1988, 1; Brown 2011, 751; and Willenken 2012, 545.

7. See Slote 1985, chap. 5; Howard-Snyder and Norcross 1993; and Norcross 2006a, 2006b.

8. Perhaps the clearest statement of this interpretation of Kant appears in Herman 1993, but it can also be found, to varying degrees, in the works of Marcia Baron, Paul Guyer, Thomas Hill Jr., Christine Korsgaard, Onora O’Neill, and Allen Wood. For discussion, see Pippin 2001, 386–87; and Ridge 2009, 421.

are, must be the sort of thing that can be promoted or prevented. For example, a particular person—call him “Bob”—cannot, strictly speaking, be promoted or prevented (in the sense of these terms intended here: we’re not talking about job promotions). So if a teleologist insists that Bob is a fundamental bearer of value, really what he or she means is that *the state of affairs in which Bob exists* has value and hence is to be promoted. For the teleologist, the true bearers of value are states of affairs,⁹ and any talk on the teleologist’s part of concrete entities (be they persons or animals, things or events, mental states or personal relationships) having fundamental value is really shorthand for talk of *the states of affairs in which those entities exist, occur, or happen* having fundamental value.¹⁰ Second, the teleologist’s conception of value commits him or her to the claim that all nonfundamental value is to be explained in terms of how well a bearer of nonfundamental value conduces toward the promotion of whatever those things are that have fundamental value; or in other words, nonfundamental value is explicable in terms of conduciveness toward fundamental value. These two constraints, together with the requirement that any facts that there might be about deontic notions such as obligation, permission, and rightness must obtain in virtue of facts about value, give us the heart of the teleological point of view.

With that in mind, we can define a teleological normative theory (whether in ethics or epistemology) as follows. Most teleological theories have three basic components: *a theory of final value*, *a theory of overall value*, and *a deontic theory*. However, a teleological theory need not have all three components; in particular, scalar versions of teleology lack a deontic theory. When all three theories are present, each successive theory crucially depends on the preceding one: the theory of overall value depends on the theory of final value, and the deontic theory depends on the theory of overall value.

A teleologist’s *theory of final value* can be formulated in two different, but equivalent, ways, depending on whether it is formulated in terms of objects/ends or in terms of goals/aims.¹¹ Formulated in terms of

9. Or entities that play roughly the same role as states of affairs, such as facts, or property instantiations, or (portions of) possible worlds.

10. Here I agree with Elizabeth Anderson (1993, 30–32) and T. M. Scanlon (1998, 80), and disagree with Douglas Portmore (2011, 129–31).

11. The more traditional name for the teleologist’s theory of final value is “a theory of intrinsic value.” However, it is now widely recognized that the intrinsic vs. extrinsic value distinction is distinct from the final vs. instrumental value distinction (see Korsgaard

objects/ends, the theory specifies a certain set of states of affairs that have value or disvalue as ends in themselves. (It might also specify the degree to which they are valuable or disvaluable in this way.) Again, if we are speaking loosely, we can talk of concrete entities such as headaches or true beliefs having value or disvalue as ends, but really it is *the state of affairs in which someone has a headache* that has final disvalue, or *the state of affairs in which someone has a true belief* that has final value. Formulated in terms of goals/aims, the teleologist's theory of final value identifies a certain list of goals or aims that structure the norms under consideration. (It might also ascribe to each of these goals or aims a weight or strength.) Since this is a theory of *final* value, these must be *ultimate* goals or aims, not goals or aims that serve other, more basic goals or aims. Some of these goals will be positive goals: a goal of bringing about some state of affairs, or of making it the case that something happens. And some of these goals will be negative goals: a goal of avoiding some state of affairs, or of preventing something from happening.¹² As should be apparent, these two ways of formulating a theory of final value are equivalent: the positive (or negative) goals in the latter formulation correspond to the ends with positive (or negative) value in the former.¹³

The teleologist's *theory of overall value* takes us from evaluations of the goodness or badness of certain states of affairs as ends in themselves to

1983). Moreover, it is the latter distinction that is most crucial for this stage of the teleologist's project: there is no reason to restrict our range of ultimate ends to those that are valuable in virtue of their intrinsic properties. For example, a theory of intrinsic value could not include true beliefs among the entities that are intrinsically valuable since other than in special cases such as a belief in the proposition $\langle I \text{ have at least one belief} \rangle$, most beliefs are true in virtue of their extrinsic properties; however, a theory of final value has no problem including true beliefs among the entities that have final value.

12. Negative goals/aims cannot be reduced to positive goals/aims: a negative goal of avoiding its being the case that p is not equivalent to a positive goal of making it the case that not- p . (To say that I have a negative goal of avoiding work is not the same as saying that I have a positive goal of not-working.) Similarly, a disinclination to ϕ is not the same as an inclination to not- ϕ , and disliking that p is not the same as liking that not- p .

13. Strictly speaking, what I here call "a theory of final value" should be "a theory of *pro tanto* final value" that focuses on the specifically *pro tanto* goodness or badness of states of affairs as ends in themselves (that is, focuses on a given state of affair's goodness or badness as an end *in a certain respect*, not its goodness or badness as an end when all of the ways in which it might be good or bad as an end are weighed against each other). Since the arguments to follow depend more on the instrumental value vs. final value distinction than on the *pro tanto* value vs. all-things-considered value distinction, I have chosen to suppress this complication in my formulation of teleological theories, thereby avoiding the need to clutter my claims with "*pro tanto*" qualifiers.

all-things-considered verdicts about the goodness or badness of any entities that conduce toward or promote those states of affairs. Because a teleologist may want to evaluatively rank many different sorts of entities—acts, beliefs, motives, rules, institutions, and so on—his or her theory of overall value might break up into several different subtheories corresponding to each of these *evaluative focal points* (to use Shelly Kagan [2000]’s helpful terminology). For each evaluative focal point, the teleologist’s theory of overall value specifies a comparative ranking of every entity that falls within that evaluative focal point in terms of how well, all things considered, it conduces toward or promotes the various states of affairs that, according to the teleologist’s theory of final value, have value or disvalue as ends in themselves. Different teleological theories will interpret what it takes for an entity X to promote or conduce toward a state of affairs S that has value or disvalue in different ways, including some or all of the following possibilities (and various combinations of these possibilities):

- X is a *causal means* to S: X’s occurrence, happening, or existence (partially or entirely) causes it to be the case that S obtains.
- X *instantiates* S: for X to occur, happen, or exist just is for S to obtain.
- X is an “*upward*” *constitutive means* to S: X’s occurrence, happening, or existence (partially or entirely) constitutes S’s obtaining.¹⁴
- X is a “*downward*” *constitutive means* to S: X’s occurrence, happening, or existence is (partially or entirely) constituted by S’s obtaining.¹⁵

Assigning an overall value to X involves (i) determining all the ways in which X stands in the relevant promoting relation to states of affairs of final value/disvalue, (ii) weighing all these various eventualities against each other, and (iii) comparing this net result to the net result for every other item that falls within X’s evaluative focal point. Depending on how precisely the teleologist’s theory of final value assigns weights or degrees

14. An example: if my having read a certain book is valuable, then my having read the first page of that book might be a way of promoting that value, in virtue of its being an upward constitutive means to my having read the entire book.

15. An example: if my intending to perform some action is valuable, then my (intentionally) performing that action might be a way of promoting that value, in virtue of its being a downward constitutive means to my having the intention in question.

of goodness, and on how precisely the teleologist specifies a weighing algorithm, the theory of overall value might result in a *cardinal ordering* (that is, a quantitative ordering that assigns specific numbers on some scale to the items being ranked), or it may merely result in an *ordinal ordering* (that is, an ordering that specifies only whether various items in the evaluative focal point have greater, lesser, or equal overall value in comparison to one another).¹⁶

The final element in a given teleological position is a *deontic theory* that assigns deontic properties such as being obligatory or permissible, being right or wrong, being justified or unjustified, on the basis of the theory of overall value.¹⁷ Again, since the teleologist may want to assign deontic properties to a variety of different entities—acts as well as beliefs, motives as well as character traits, rules as well as institutions—his or her deontic theory might split up into various subtheories corresponding to each of these *deontic focal points*. These subtheories can take two forms. If the deontic focal point is also an evaluative focal point, then the deontic subtheory will be a *direct theory* that assigns a deontic status to the items in the deontic focal point based on their ranking according to the portion of the teleologist's theory of overall value devoted to that evaluative focal point. There are various ways that this assignment could proceed: it could be a *maximizing theory* according to which item X in deontic focal point D is right if and only if there is no other item in D that has greater overall value than X; it could be a *satisficing theory* according to which item X in deontic focal point D is right if and only if its overall value is above some threshold; or the assignment could depend on the theory of overall value in a more complex manner.¹⁸ If the deontic focal point is not also an evaluative focal point, then the deontic subtheory for that focal point will be an *indirect theory* that assigns deontic properties to the members of that deontic focal point on the basis of a bridge principle linking that deontic focal point to another deontic focal point for which deontic statuses have already been assigned. The most familiar indirect deontic subtheory is the sort posited by rule-consequentialism, in which right acts are defined

16. Note that if some of the teleologist's final values are incommensurable, incomparable, or vague, the resulting overall ordinal ordering might not be a total ordering (that is, there might be two items in the evaluative focal point such that the first is neither greater than, less than, nor equal to the second in overall value).

17. I assume here that *being justified* is a deontic property. If you disagree, replace the phrase 'deontic theory' with a term that encompasses your desired way of categorizing facts about justification.

18. See, for example, Sider 1993 and Portmore 2008.

in terms of right rules, and the rightness of rules is directly assessed in terms of how well (when properly internalized by a suitable portion of the population) the rules promote the good and don't promote the bad.¹⁹

My characterization of teleological normative theories has, of necessity, been quite abstract, but some examples will help bring our discussion down to earth. First, two examples of teleological theories in ethics:

Maximizing hedonistic act-utilitarianism:

- i. *Theory of final value:*
 Pleasurable experiences have value as ends.
 Painful experiences have disvalue as ends.
 Nothing else has value or disvalue as an end.
- ii. *Theory of overall value* (for evaluative focal point {acts}):
 S's ϕ -ing at time t has more overall value than S's ψ -ing at t iff [the net balance of pleasure over pain that would be brought about if S ϕ 's at t] > [the net balance of pleasure over pain that would be brought about if S ψ 's at t].
- iii. *Deontic theory* (for deontic focal point {acts} and deontic property *being right*):
 S's ϕ -ing at time t is right iff no other act available to S at t has more overall value.

*Satisficing hedonistic motive-utilitarianism:*²⁰

- i. *Theory of final value:*
 Pleasurable experiences have value as ends.
 Painful experiences have disvalue as ends.
 Nothing else has value or disvalue as an end.
- ii. *Theory of overall value* (for evaluative focal point {motive-sets}):
 Motive-set m has more overall value than motive-set m' iff m tends (when possessed in a suitable range of circum-

19. Different versions of rule-consequentialism vary, of course, in how they specify the states of affairs promoted by a given rule (or system of rules) and in how they link the deontic properties of rules to the deontic properties of acts. For some variations, see Pettit and Smith 2000, 130–31; and Kagan 2000, 137–38.

20. See Adams 1976, 478–79, although I have changed Adams's proposal in several ways (for instance, switching it from a maximizing to a satisficing theory) to make this example more analogous to an example from the epistemic realm that I discuss momentarily.

stances) to lead to a greater balance of pleasure over pain in the world than m' does.

- iii. *Deontic theory* (for deontic focal point {motive-sets} and deontic property *being right*):

Motive-set m is right iff [m 's overall value $\geq$ threshold T].

Deontic theory (for deontic focal point {acts} and deontic property *being right*):

S's ϕ -ing at time t is right iff the motive-set that caused S to ϕ at t is right.

Both of these normative theories are examples of ethical teleologies: they specify certain ultimate ends that we, as agents, should promote and then assign deontic properties on the basis of how well acts, motive-sets, and the like conduce (whether directly or indirectly) toward those ends. In order to provide an example of an epistemic teleology, we need to focus on items that have distinctively *epistemic* value. In other words, we need to specify certain ultimate *epistemic* ends that we, as believers, should promote and then assign deontic properties on the basis of how well beliefs, cognitive processes, habits of thought, and the like conduce (whether directly or indirectly) toward those ends. Perhaps the most famous example of a teleological epistemic theory of this sort is process reliabilism:

(Simplified) process reliabilism:²¹

- i. *Theory of final value*:

True beliefs have epistemic value as ends.

False beliefs have epistemic disvalue as ends.

Nothing else has epistemic value or disvalue as an end.

- ii. *Theory of overall value* (for evaluative focal point {belief-forming processes}):

Belief-forming process b has more overall epistemic value than belief-forming process b' iff b tends (when employed in a suitable range of circumstances) to yield a greater ratio of true to false beliefs than b' does.

- iii. *Deontic theory* (for deontic focal point {belief-forming processes} and deontic property *being reliable*):

21. See Goldman 1992 [1979]. I add the qualification 'simplified' since this formulation of process reliabilism ignores the distinction between belief-dependent and belief-independent cognitive processes, which for our purposes we can disregard.

Belief-forming process b is reliable iff [b 's overall epistemic value $\geq$ threshold T].

Deontic theory (for deontic focal point {beliefs} and deontic property *being justified*):

S's belief that p at time t is justified iff the belief-forming process that caused S to believe that p at t is reliable.

When observed side-by-side, the similarities between satisficing hedonistic motive-utilitarianism and process reliabilism should be quite striking. According to the former, an act is right iff it is caused by a motive-set that tends to do sufficiently well at bringing about one end (pleasurable experiences) and avoiding another end (painful experiences). According to the latter, a belief is justified iff it is caused by a process that tends to do sufficiently well at bringing about one end (true beliefs) and avoiding another end (false beliefs). These similarities are, I believe, no mere coincidence. Just as utilitarianism is the paradigm example of a teleological ethical theory, so too, I believe, reliabilism is the paradigm example of a teleological epistemic theory. This is not hard to see when we note that a more perspicuous—though far uglier—name for reliabilism is “truth-conducivism”: *process reliabilists* hold that a belief is justified when it is caused by a sufficiently truth-conducive process, *indicator reliabilists* hold that a belief is justified when it is connected in some way to a sufficiently truth-conducive indicator, *method reliabilists* hold that a belief is justified when it is formed on the basis of a sufficiently truth-conducive method of belief formation, and so on. Thus if we take our sole ultimate epistemic end to be truth-in-belief (plus avoidance of falsity-in-belief), and if we make the distinctively teleological assumption that every other normative epistemic property is to be explained in terms of how well the thing bearing that property conduces toward the fulfillment of our ultimate epistemic ends, it is all too easy to slide into defending a version of reliabilism.

3. The Pervasiveness of Teleology in Contemporary Epistemology

I have just identified a widely discussed epistemic theory—namely, reliabilism—that counts as teleological in structure. Once the basic distinction between teleological and nonteleological epistemic theories is pointed out, one naturally expects that, just as in the ethics literature where a wide variety of teleological theories are defended and a wide variety of nonteleological theories are defended, so too the epistemology literature should be populated by a wide variety of both teleological and nonteleological theories. But what one finds when one turns to the

epistemology literature as it exists today is quite surprising, for a teleological approach to normative epistemology is overwhelmingly the dominant view.

Over the past three decades, introductory epistemology courses have tended to have their own familiar narrative: namely, the ongoing struggle between internalist and externalist theories of justification. Internalists about justification hold that the facts that determine whether a belief of mine is justified must always be accessible to me upon reflection, or must always be facts about my own (nonfactive) mental states. Externalists about justification deny this: they hold that whether or not I am justified in believing something can, at least sometimes, be determined by facts that are beyond my ken, or by facts that are not entirely about my own (nonfactive) mental states.²² Because so much work in epistemology over the past few decades has centered on the conflict between internalists and externalists, structuring an introductory epistemology course around the clash between these two camps is difficult to resist. But what I, at least, find eye-opening is that when one considers the major figures in the debate between internalists and externalists about epistemic justification (thought by many to be *the* central debate in contemporary epistemology), almost all of them are committed teleologists about epistemic normativity.

Consider, first, externalists about epistemic justification. The canonical externalist theory is reliabilism, and the foremost defender of reliabilism is Alvin Goldman. Over the years, Goldman has been quite frank about his adherence to a teleological conception of epistemic normativity. In his first monograph devoted to defending reliabilism, Goldman (1986, 97) explicitly endorses a “consequentialist” rather than a “deontological” criterion for the rightness of a given system of justificational rules. In particular, he endorses a criterion whereby the consequences that matter are one’s believing truths and not believing falsehoods (Goldman 1986, 98). In a more recent article, Goldman (2001, 32) writes that “true belief is the ultimate value in the epistemic sphere, and various belief-forming processes, faculties, or mechanisms

22. I have characterized internalism and externalism disjunctively, in order to sidestep a current debate over whether internalism about epistemic justification is best formulated as what Earl Conee and Richard Feldman call “accessibilism” (where the internal is construed as that-to-which-I-have-some-special-sort-of-access) or as what they call “mentalism” (where the internal is construed as that-which-is-internal-to-my-mental-life). See Conee and Feldman 2004 [2001], 55. William Alston (1989 [1986]) makes a similar distinction between what he calls “access internalism” and “perspectival internalism.”

are licensed as [epistemically] virtuous because they are conducive to true belief. Beliefs are justified when they are produced by these very truth-conducive processes." Thus, on his account, "The principal relation that epistemic virtues bear to the core epistemic value will be a teleological or consequentialist one. A process, trait, or action is an epistemic virtue to the extent that it tends to produce, generate, or promote (roughly) true belief" (Goldman 2001, 31). Goldman's commitment to a teleological/consequentialist normative framework could not be any more explicit, and many other externalists follow him in this regard.

So let us turn to internalists about epistemic justification. Two of the foremost defenders of internalist theories over the years have been Laurence Bonjour and Richard Foley. Although in his earlier work Bonjour defended coherentism about a posteriori justification (Bonjour 1976, 1978, 1985) and later defected, becoming a thorough-going foundationalist (Bonjour 1999a, 1999b, 2003), all along Bonjour has defended internalist versions of these epistemological positions and has vigorously argued against externalist approaches to epistemic norms (Bonjour 1980, 2002). Given this, one might expect Bonjour to be opposed to Goldman's teleological/consequentialist outlook on normative epistemology. But far from being opposed to such an outlook, Bonjour is in fact one of its chief proponents. In a typical passage, Bonjour (1985, 7–8) writes,

What makes us cognitive beings at all is our capacity for belief, and the goal of our distinctively cognitive endeavors is *truth*: we want our beliefs to correctly and accurately depict the world. . . . The basic role of justification is that of a *means* to truth, a more directly attainable mediating link between our subjective starting point and our objective goal. . . . If epistemic justification were not conducive to truth in this way, if finding epistemically justified beliefs did not substantially increase the likelihood of finding true ones, then epistemic justification would be irrelevant to our main cognitive goal and of dubious worth. . . . Epistemic justification is therefore in the final analysis only an instrumental value, not an intrinsic one.

This essentially teleological way of thinking about the nature of epistemic justification is not an incidental part of Bonjour's epistemological program. Not only does he affirm, time and time again,²³ this idea that the

23. See, among other places, Bonjour 1976, 289; 1978, 5; 1980, 54; 1986, 94; 1998, 1; and 2010, 35.

features that render a belief epistemically justified must conduce toward its truth in a subjectively accessible manner, but moreover this idea serves as a crucial premise in many of BonJour's most famous arguments and objections, including (a) his old argument against foundationalism from his coherentist days (BonJour 1976, sec. 2; 1978, sec. 2; 1985, sec. 2.3), (b) the objection to coherentism that most worried him when he was a coherentist (BonJour 1976, sec. 4; 1985, sec. 5.5), (c) his argument for including one of the more controversial elements in his version of coherentism, namely the "doxastic presumption" that a person's metarepresentation of the contents of his or her entire system of beliefs is roughly correct (BonJour 1985, sec. 5.4; 2003, sec. 3.1.4), and (d) one of the objections to coherentism that eventually convinced him of its falsity (BonJour 1999a, 130; 2003, 58–59). Suffice to say, BonJour's commitment to epistemic teleology runs deep.²⁴

We find much the same thing with Richard Foley. Over the years Foley has defended what is widely regarded as the most thoroughly internalistic account of epistemic rationality yet developed. In fact, Foley's internalist conception of epistemic rationality has become so well known that it is usually referred to with his name attached to it: "Foley-rationality." To first approximation, a belief is Foley-rational iff it would survive a Cartesian meditation. To second approximation, a belief is Foley-rational iff it is the conclusion of an argument whose premises and inferences the subject would find uncontroversial if he or she were appropriately reflective and had unlimited time to reach a stable point of view.²⁵ If this is all one knew about Foley's view, one might expect him to justify it with a nonteleological conception of epistemic normativity, especially given the Cartesian nature of his program. As it turns out, though, Foley is an unabashed teleologist. Both of Foley's books defending his account of epistemic rationality (Foley 1987 and 1993) open with chapters insisting that rationality in general is goal-oriented and that epistemic rationality in particular is structured around the fundamental goal of "now believing those propositions that are true and now not

24. Sometimes BonJour (1985, 8) glosses the relation between epistemic justification and the cognitive goal of truth in a way that is not distinctively teleological, as amounting to little more than the requirement that "one accept all and only those beliefs which one has good reason to think are true." However, this more minimal claim is *not* equivalent to the requirement of subjectively accessible truth-conduciveness. Moreover, it is the latter requirement that plays a role in (a) through (d).

25. For further approximations, see Foley 1987, 1993.

believing those propositions that are false” (Foley 1993, 19). Though an internalist, Foley is also a card-carrying teleologist.

Finally, let us consider a figure who is one of the foremost defenders of a hybrid epistemic theory that incorporates elements from both the externalist and internalist traditions, namely, William Alston. In his most recent book, *Beyond “Justification”: Dimensions of Epistemic Evaluation*, Alston resists the idea that there is only one sort of positive epistemic status—call it “justification”—that a belief might possess. Instead, he defends a pluralist theory according to which beliefs can have a variety of positive epistemic statuses—or “epistemic desiderata,” as he calls them—some of which are more externalist in nature (for example, reliability), others of which are more internalist (for example, adequacy of grounds of belief). Thus it might seem that Alston’s pluralist approach to normative epistemology would allow him to countenance epistemic desiderata not tied in a teleological/consequentialist way to the furthering of some goal or end, just as W. D. Ross’s pluralist position in normative ethics includes both consequentialist and nonconsequentialist elements. However, when Alston (2005, 29) turns to explaining why each of the desiderata he considers counts as an *epistemic* desideratum (rather than, say, a *moral* or *prudential* desideratum), he restricts himself to explanations that are decidedly teleological in character:

Epistemology consists of a critical reflection on human cognition. And the evaluative aspect of epistemology involves an attempt to identify ways in which the conduct and the products of our cognitive activities can be better or worse vis-à-vis the goals of cognition. And what are those goals? Along with many other epistemologists I suggest that the primary function of cognition in human life is to acquire true beliefs rather than false beliefs about matters that are of interest or importance to us.

Thus, for Alston, all epistemic desiderata are ultimately defined in terms of how well they help us further “the goals of cognition.” A little later, Alston (2005, 32) makes it clear that he embraces a maximizing conception of these goals: our fundamental epistemic goal, he insists, is not just to acquire true rather than false beliefs about matters that are of interest or importance to us, but moreover to maximize our number of true beliefs about such matters and to minimize our number of false beliefs. This assumption of a teleological framework for epistemological theorizing is not a recent development in Alston’s thought. In countless places in his seminal earlier work on the internalism/externalism debate, on doxastic voluntarism, on epistemic circularity, and on the

epistemology of religious experience, Alston restates—and relies upon—his view that epistemic evaluation is always evaluation with regard to our basic cognitive goal of maximizing the number of one's true beliefs and minimizing the number of one's false beliefs in some body of belief.²⁶

Thus we see that one major externalist (Goldman), two major internalists (BonJour, Foley), and one major compromiser between externalism and internalism (Alston) are all committed teleologists. These examples have not been cherry-picked to suit my purposes: the number of epistemologists from across the externalist-internalist spectrum who adhere to a basically teleological perspective on epistemic normativity is astounding.²⁷ There are, of course, a few exceptions.²⁸ But

26. See, among other places, Alston 1989 [1976], 305; 1989 [1985], 83; 1989 [1988a], 116; 1989 [1988b], 231–32; 1989, 3; 1991, 72; and 1993, 3–4. (In most of these earlier works, however, Alston does not add the proviso that the beliefs in question must be ones “of interest or importance to us.”) Note that not only does Alston restate his commitment to teleology in all of these places, but moreover that commitment almost always serves as a premise in the arguments that follow. To give but one example: during a crucial point in his well-known argument against what he calls “the deontological conception of epistemic justification,” Alston (1989 [1988a], 143–52) argues that certain beliefs cannot be justified since they were not formed in a truth-conducive way (nor in a way the subject is justified in believing to be truth conducive).

27. Other adherents to a teleological approach to epistemic normativity include, but are not limited to, Ralph Baergen (1995), Anthony Booth (2006, 2008), Berit Brogaard (2009), Marian David (2001, 2005), Brian Ellis (1988), Richard Feldman (1988, 2002), Hartry Field (1982, 1998, 2000), Peter J. Graham (2011), John Greco (1993, 1999), Stephen Grimm (2008, 2009), Richard Hall and Charles Johnson (1998), Allan Hazlett (2006, n.d.), Hilary Kornblith (1985, 1993, 2002), Jonathan Kvanvig (2003, 2005, 2011), Keith Lehrer (1977, 1981, 1988), Jarrett Leplin (2007, 2009), Michael Lynch (2004, 2009a, 2009b), Stephen Maitzen (1995), Nenad Miscevic (2007), James Montmarquet (1987, 1993, 2008), Philip Percival (2002), Alvin Plantinga (1988, 1993a, 1993b), Steven Reynolds (2002), Wayne Riggs (2003, 2008a, 2008b), Darrell Rowbottom (2008), Frederick Schmitt (1992, 2000), Tomoji Shogenji (2012a, 2012b), Asbjørn Steglich-Petersen (2009, 2011, n.d.), Hamid Vahid (2006, 2010, 2011), Ralph Wedgwood (2002, 2008), Jonathan Weinberg (2007), Dennis Whitcomb (2007, 2011), and at least certain time-slices of Paul Moser (1985) and Ernest Sosa (1991).

One issue I do not take a stand on in the current essay is whether defenders of sensitivity/safety conditions on knowledge/justification should be regarded as epistemic teleologists, in my sense. Settling whether satisfying one of these conditions counts as a way of conducing toward or promoting epistemic value would take us too far afield, and extending the arguments to come so that they encompass sensitivity and safety views would make an already complicated argument hopelessly byzantine. I hope to address the topic of whether sensitivity and safety views are susceptible to variants of the arguments found in this essay in future work.

28. For example, Richard Fumerton (2001, 54–56; 2002, 208–9; 2006, 34–35)

overwhelmingly, much of the recent debate between internalists and externalists in epistemology can be seen as an in-house dispute among teleologists.

During the latter portion of the twentieth century, this basic assumption of a teleological approach to normative epistemology was largely under the surface. But more recently, it has become its own area of research. There is now an active debate over what “our epistemic goals” are: whether they are just the attainment of true beliefs and the avoidance of false beliefs, as some contend,²⁹ or whether this list needs to be revised or expanded to include other items.³⁰ And in the past decade there have been conferences, workshops, and blogs devoted exclusively to the topics of “epistemic value” and “epistemic goodness.” One author has even gone so far as to proclaim that we are undergoing a “value turn in epistemology” as more and more epistemologists recognize the importance of what he calls “value-driven epistemology”: an approach to normative epistemology that proceeds by first specifying our basic epistemic goals or values and then connecting all other normative notions in epistemology to the furtherance of these goals or the promotion of these values (Riggs 2008b).

Lost in the shuffle during this ferment of excitement about epistemic goals, epistemic value, value-driven epistemology, and the like has been the possibility that we might *not* have distinctively epistemic goals, that epistemic norms might *not* be ultimately grounded in facts about epistemic value, and that the teleological/consequentialist commitments of value-driven epistemology might *not* be correct. My favorite example of the overlooking of such a possibility occurs in the recent volume *Contemporary Debates in Epistemology*, edited by Matthias Steup and Ernest Sosa. In it, a number of classic epistemological questions such as “Can skepticism be refuted?” and “Is there a priori knowledge?” are debated by a group of prominent epistemologists. For each question, one author defends the “yes” answer to the question, and another author defends the “no” answer. One of these questions is “Is truth the primary epistemic goal?” with Marian David (2005) arguing that it is and Jonathan Kvanvig (2005)

argues against construing claims about what we ought to believe as claims about the efficacy of achieving certain goals or ends. Note, also, that quite a few epistemologists are noncommittal in their work about the truth or falsity of a teleological approach to epistemology.

29. Most prominently, Marian David (2001, 2005).

30. As Stephen Maitzen (1995), Michael DePaul (2001), Wayne Riggs (2002), Jonathan Kvanvig (2005), Berit Brogaard (2009), and Stephen Grimm (2009) have all argued.

arguing that there are other primary epistemic goals beyond truth. But the question “Is truth the primary epistemic goal?” of course presupposes that we have epistemic goals, some of which are more primary than others, and neither David nor Kvanvig disputes this. Somehow during the past few decades a teleological approach to epistemic normativity has become so uncontroversial that it is now not even up for debate.

4. Ethical Trade-Offs and the Separateness of Persons

My aim in this essay, as I have already made clear, is to argue against this growing trend of assuming that epistemic theories must be teleological in structure. I think that teleology/consequentialism in normative ethics has its attractions, but these attractions do not all translate well when applied to normative epistemology. It is one thing to have a goal-oriented approach to what we should do—to ground facts about should-be-doneness in facts about what conduces toward the promotion of value. But a goal-oriented approach to what we should believe? Such a conception of epistemology, I believe, yields a variety of implausible results. Facts about what we should believe do not, I believe, obtain in virtue of facts about what conduces toward the promotion of items that possess a distinctively epistemic sort of value. Moreover, if they did, then teleological theories in epistemology would look a lot more like teleological theories in ethics than any epistemologist is—or should be—willing to accept. Or so, at least, I will argue. But first I need to quickly review some familiar features of teleological/consequentialist theories in ethics, since they will prove important in the argument to come.

One of the greatest virtues, as well as one of the greatest vices, of the teleological/consequentialist approach to ethics is its countenancing of trade-offs in its determination of what we should do. Most teleological ethical theories have no problem explaining why we should sometimes sacrifice one person’s life in order to save the lives of five people, or why we should sometimes allow a certain tragedy to befall several people in order to prevent an even greater tragedy from befalling a much larger group of people. I call this feature of teleological ethical theories a virtue because, in a large number of cases, allowing or prescribing such a trade-off strikes most of us as the correct result, and teleological theories can easily explain why this might be so. Most nonteleological ethical theories, on the other hand, have a more difficult time accommodating trade-offs. This leads to what I consider to be one of the deepest challenges for nonteleological/nonconsequentialist theories in ethics: how to include

an aggregative element in one's theory, so as to allow for trade-offs in some cases, without that aggregative element overrunning one's theory and turning it into a version of teleology/consequentialism. How does one make the numbers sometimes count, and sometimes carry the day, without making them always count, and always carry the day? This is a problem that many who adopt a nonteleological approach to normative ethics must grapple with.³¹

However, this countenancing of trade-offs is also, notoriously, one of ethical teleology's greatest vices. The ease with which teleological ethical theories allow—and usually mandate—trade-offs is what gives rise to many of the most famous intuitive counterexamples to such theories: think, here, of cases in which a doctor has the ability to trade the life of a healthy individual attending a routine check-up for the lives of five people each in need of a new organ (Thomson 1976, 206), or in which a sheriff has the opportunity to trade the framing of an innocent man for the health and happiness of those who would otherwise be harmed by an angry mob (McCloskey 1957, 468–69), or in which a botanist has the chance to trade the murder of one South American villager by his own hand for the murder of twenty villagers by the hand of a soldier from a military junta (Williams 1973b, 98–99). When the trade-offs in question are interpersonal—that is, when the trade-offs involve harming or hurting one person in order to benefit or advantage a second person—then the common refrain is that, in permitting or requiring such trade-offs, teleological ethical theories “ignore the separateness of persons.”³² The idea here is that teleological theories such as act-utilitarianism do not treat *intrapersonal* trade-offs—trade-offs that involve harming or hurting a given person in one way in order to benefit or advantage that same person in another way—as fundamentally any different from *interpersonal* trade-offs when determining what an agent should do, but whereas intrapersonal trade-offs are morally defensible when the benefit to the one person outweighs the harm, interpersonal trade-offs are not morally defensible in the same way, or at least not always morally defensible in the same way.

31. Consider, for example, T. M. Scanlon's (1998, 229–41) attempt to incorporate a limited aggregative element in his contractualism in *What We Owe to Each Other*, which Scanlon (2002, 354) later singles out as the part of that book with which he is the least satisfied.

32. A worry most famously voiced by John Rawls (1963, 124; 1971, 26–27) but also endorsed, in different forms, by David Gauthier (1962, 126), Thomas Nagel (1970, 134), Robert Nozick (1974, 32–33), and a host of others.

I take no stand here on whether those teleological ethical theories that ignore the separateness of persons are making a moral mistake in doing so. My point, rather, is that it is ethical teleology's proclivity for trade-offs that opens some (but not all) versions of it to the charge of ignoring people's separateness. Ethical teleologists tend to respond to this charge in one of two ways: either they attempt to reformulate their theory so that it does not deem the problematic sorts of interpersonal trade-offs to be morally permissible, or they argue that—common sense be damned—the allegedly untenable interpersonal trade-offs allowed by their theory are in fact morally defensible.

Why, though, are teleological ethical theories so amenable to trade-offs, whether inter- or intrapersonal? It is, I believe, because of a deep structural feature possessed by all teleological ethical theories, together with two more specific features possessed by many (but not all) such theories. The deep structural feature is this: all teleological ethical theories build a theory of overall value out of a theory of final value. The first more specific feature is this: most teleologists defend a theory of final value that allows for the possibility of conflicts or competition between the values making up that theory. Put these two features together, and we have a problem: in cases of conflict or competition between final values, how do we determine overall value? For example, what do we do if an item in a given evaluative focal point conduces toward a state of affairs that is good as an end but also conduces toward a different state of affairs that is bad as an end (a case of conflict), or if one item in an evaluative focal point promotes a state of affairs that is good as an end, and another item in the same evaluative focal point promotes a different state of affairs that is also good as an end (a case of competition)? In order to reach a fixed verdict about overall value, it seems that the teleologist needs some way of trading off these values against one another, some way of (to use another common metaphor) weighing the final goodness/badness of one eventuality against the final goodness/badness of another. The most common approach among teleologists for settling these matters is through aggregation: all of the values and disvalues promoted by a given item are somehow aggregated together (either additively or otherwise), and the conglomerated sum is compared to the conglomerated sum that we get for every other item in the item's evaluative focal point. This is the second more specific feature that explains why many teleological ethical theories are so amenable to trade-offs is this: when conflicts and competition among final values are settled through aggregation, the allowing or prescribing of trade-offs becomes rampant.

I have just identified one deep structural feature possessed by all teleological ethical theories plus two more specific features possessed by many such theories that together, I claim, explain why ethical teleology so often countenances a wide variety of trade-offs. But now here's the crucial point: many, if not all, teleological *epistemic* theories share these three features. They, too, build a theory of overall value out of a theory of final value. They, too, allow for conflicts and competition between items of final value. And they, too, settle these conflicts and competition through aggregation. So we should expect most teleological epistemic theories to countenance a wide variety of trade-offs. But that's precisely where they run into trouble.

5. Epistemic Trade-Offs and the Separateness of Propositions

The time has finally come to argue against the teleological approach to normative epistemology. I aim my argument to be as general as possible, but it will help to start by constructing an argument against a slightly narrower target and then generalizing later.

So let us begin by considering a version of epistemic teleology that holds the following theory of final epistemic value: true beliefs have epistemic value as ends, false beliefs have epistemic disvalue as ends, and nothing else has epistemic value or disvalue as an end. Or to formulate things in terms of goals instead of ends: according to this theory of final epistemic value, our only ultimate epistemic goals are (i) the accumulation of true beliefs and (ii) the avoidance of false beliefs. Adapting some terminology of Goldman's (Goldman 1999, 5; 2001, 31), let us refer to this theory of final epistemic value as "veritism" and let us call teleological epistemic theories that include veritism as their first component "veritistic."

Veritism is, in many ways, the default theory of final value for epistemic teleologists. Indeed, in the quotations I provided in section 3, Goldman, Bonjour, Foley, and Alston all endorse versions of veritism. However, the attentive reader will note that although Foley and Alston refer to both the positive goal of acquiring true beliefs and the negative goal of avoiding false beliefs, Goldman and Bonjour only mention the first of these. Why, then, include two goals rather than one?

That we need two separate goals here is a point usually attributed to William James (1897, 17–19). If we include only a positive goal of accumulating true beliefs, that would encourage *epistemic recklessness*. In the limit, one could trivially satisfy this goal by believing every proposition; of course, doing so would ensure that one believes a large

number of falsehoods since for every proposition one believes, one will also believe its negation, but if there is no negative goal of avoiding false belief, then no epistemic harm is done in believing a few extra falsehoods. On the other hand, if we include in our theory of final value a negative goal of avoiding false beliefs without a corresponding positive goal of acquiring true beliefs, that would encourage *undue epistemic caution*. In the limit, one could trivially satisfy the negative goal by not believing anything, or—if that is not possible—by believing as little as one can. In James's (1897, 18) memorable words, someone who advocates the avoidance of error without also stressing the importance of believing the truth "merely shows his own preponderant private horror of becoming a dupe." For these Jamesian reasons, it is widely (though not universally: see Piller 2009) accepted that we need *both* of these goals to capture what people are after when they talk of "the search for truth" or say things like "I want the truth." For convenience, however, many authors who endorse these twin goals group them together under a single heading that they call "the truth-goal,"³³ and we can assume that Goldman and BonJour are following this practice in the passages quoted earlier.

Three points of clarification before I proceed:

1. Veritism is logically distinct from the view, first proposed by Bernard Williams (1973a [1970], 136–37) but since defended by others (including David Velleman [2000a], Ralph Wedgwood [2002], Pamela Hieronymi [2006], and Asbjørn Steglich-Petersen [2009]), that it is a constitutive feature of beliefs that they "aim at the truth." Appealing to such a view about the nature of belief is one natural way of motivating veritism. However, it is not the only way. Compare: one need not hold that actions constitutively aim toward pleasure and away from pain in order to hold the hedonistic theory that pleasure and the avoidance of pain are our ultimate practical ends. (Hedonistic act-utilitarians are not logically committed to the thesis that action has a constitutive aim.) And, conversely, those who hold that action has a constitutive aim are not required to be consequentialists; indeed, so-called constitutivists about practical reason rarely are consequentialists (see Korsgaard 2008, 2009; Velleman 2000b, 2009). So, too, in the epistemic case: veritism neither entails,

33. See, for instance, David 2001, 152–53.

- nor is entailed by, the view that it is a constitutive feature of beliefs that they “aim at the truth.”
2. Although it is common for defenders of veritism to speak of the acquisition of true beliefs and the avoidance of false beliefs as being “our” fundamental epistemic goals, epistemic teleology—as I understand it here—is not committed to the view that every agent as a matter of fact possesses the goals featured in its theory of final epistemic value. Thus it is no objection to veritism that some of us don’t desire to believe the truth about every proposition. Again, a comparison with teleological ethical theories that include a hedonistic theory of final value is instructive: it is no objection to hedonistic act-utilitarianism that there are some pleasurable experiences that some of us don’t desire to bring about, or that there are some painful experiences that some of us are not averse to bringing about. Rather, what the hedonistic act-utilitarian is committed to is the following: all ethical evaluation is ultimately evaluation with regard to the twin goals of promoting pleasurable experiences and preventing painful experiences. Similarly, an epistemic teleologist who adheres to veritism is committed to the following: all epistemic evaluation is ultimately evaluation with regard to the twin goals of acquiring true beliefs and avoiding false beliefs, whether or not the agent being evaluated possesses these goals.³⁴
 3. I have been speaking here as if veritism countenances just two ultimate epistemic goals, but strictly speaking that is not accurate. Rather, what we have is a vast plethora of positive ultimate epistemic ends (to believe $\langle p \rangle$ if $\langle p \rangle$ is true, to believe $\langle q \rangle$ if $\langle q \rangle$ is true, to believe $\langle r \rangle$ if $\langle r \rangle$ is true, and so on, for every proposition), and a vast plethora of negative ultimate epistemic ends (to not believe $\langle p \rangle$ if $\langle p \rangle$ is false, to not believe $\langle q \rangle$ if $\langle q \rangle$ is false, to

34. Thus the target of my argument in this essay is importantly different from the target of Kelly 2003. Thomas Kelly is criticizing a view he calls “the instrumental conception of epistemic rationality,” but as Kelly understands that view, it holds that epistemic rationality is instrumental rationality in the service of epistemic goals that we actually possess, and Kelly’s argument crucially turns on denying that we always have the epistemic goals in question.

not believe $\langle r \rangle$ if $\langle r \rangle$ is false, and so on, for every proposition).³⁵ Some writers object to this way of conceiving of the commitments of veritism on the grounds that it implausibly ascribes to agents an infinite number of goals (David 2001, 165), but in light of the previous point, we can see that this objection is neither here nor there: to say that there are an infinite number of ultimate goals structuring epistemic norms is not the same as saying that every agent who is subject to those norms must possess an infinite number of goals.³⁶

This last comment points the way to the central problem with veritistic teleological theories. I said earlier that we should expect most epistemic teleological theories to permit or prescribe trade-offs between their fundamental aims. That veritistic teleological theories need to allow for trade-offs between the positive goal of believing a proposition if it is true and the negative goal of not believing that same proposition if it is false is a familiar point in the literature on epistemic goals.³⁷ But what has not, in my opinion, been sufficiently appreciated is the degree to which veritistic theories also need to allow for trade-offs between our epistemic goals *with respect to different propositions*: they must balance the positive goal of believing one proposition if it is true against the positive and negative goals associated with believing all other propositions, and they must balance the negative goal of not believing one proposition if it is false against the positive and negative goals associated with believing all other propositions. But such trade-offs, I put forward, are beyond the pale.

To see why, consider the following case, which is originally due to Roderick Firth (1998 [1978], 326; 1998 [1981], 261), although the version of it that I am using I owe to Richard Fumerton (2001, 55).³⁸ Suppose

35. Here and throughout, I follow the convention of using " $\langle p \rangle$ " as a name for the proposition that p .

36. Moreover, it is not clear to me what would be so objectionable about ascribing an infinite number of goals to an ordinary agent, especially if one holds a broadly dispositionalist account of belief and desire (such as functionalism). After all, most finite objects possess an infinite number of dispositions. So why, then, can't those who hold a dispositionalist account of belief and desire say that finite agents are able to possess an infinite number of beliefs and desires, and hence an infinite number of goals?

37. For example, Wayne Riggs (2003, 2008a) makes much of this point.

38. Similar uses of similar cases can be found in Blumenfeld and Blumenfeld 1978, 253; Booth 2006, 137–38; Christensen 2004, 4–5 and 173; Conee 2004 [1992], 249–51 and 256–57; Feldman 1988, 249–50; Feldman 2008, 346; Foley 1987, 8; Foley 1993, 20;

I am a scientist seeking to get a grant from a religious organization. Suppose, also, that I am an atheist: I have thought long and hard about whether God exists and have eventually come to the conclusion that He does not. However, I realize that my only chance of receiving funding from the organization is to believe in the existence of God: they only give grants to believers, and I know I am such a bad liar that I won't be able to convince the organization's review board that I believe God exists unless I genuinely do. Finally, I know that, were I to receive the grant, I would use it to further my research, which would allow me to form a large number of new true beliefs and to revise a large number of previously held false beliefs about a variety of matters of great intellectual significance. Given these circumstances, should I form a belief that God exists? Would such a belief be epistemically rational, or reasonable, or justified?

I think most of us will agree that were I to form a belief in God's existence on this basis, I would not be epistemically rational or reasonable or justified in doing so. And this is despite the fact that, in so believing, I would be greatly furthering what veritism deems to be my ultimate epistemic ends, namely, the accumulation of true beliefs and the avoidance of false beliefs. From veritism's perspective, my coming to believe in God's existence would be a sacrifice for the greater (epistemic) good: I would be trading off the pursuit of one thing of final epistemic value (true belief and the avoidance of false belief with regard to the proposition <God exists>) in order to greatly facilitate my pursuit of a variety of other things of final epistemic value (true belief and the avoidance of false belief with regard to the propositions that make up the subject matter of my scientific research). It is difficult to see how veritism—at least as we have defined it so far—can avoid saying that such a trade-off would maximize overall epistemic value. And this would seem to be so regardless of whether the veritist's theory of overall epistemic value is applied to the evaluative focal point {beliefs} or to other evaluative focal points such as {belief-forming processes} or {methods of belief formation}: a belief-forming process that yields a belief in God's existence in the given circumstance would, surely, tend to promote veritism's ultimate epistemic ends better than one that doesn't, and similarly for a method of

and Jenkins 2007, 37. However, all of these authors use these examples only to argue against extremely crude forms of epistemic teleology. One of the central tasks of the current essay is to show that Firth-style examples are a problem even for very complicated and sophisticated teleological epistemic theories.

belief formation that results in such a belief in such situations.³⁹ So regardless of how we fill in the veritist's theory of overall value and deontic theory in order to form a complete teleological epistemic theory, it appears that the veritist is committed to the highly implausible claim that my coming to believe in God's existence in order to get the grant is epistemically rational/reasonable/justified.

My aim here is not to argue by counterexample. Rather, I want to use this example to illustrate a more general point that I think is undoubtedly correct, even if, because of certain idiosyncratic details, the current example is not the best way to illustrate it. The more general point is this: when determining the epistemic status of a belief in a given proposition, it is epistemically irrelevant whether or not that belief conduces (either directly or indirectly) toward the promotion of true belief and the avoidance of false belief in *other* propositions beyond the one in question. Veritistic epistemic teleology, in flouting this fact, ignores what we might call "the epistemic separateness of propositions," just as many forms of consequentialism ignore "the ethical separateness of persons."⁴⁰ Moreover, whereas in the ethical case there is some room for arguing that it is not a normative mistake to ignore the separateness of persons, since all of us think that at least some trade-offs that cross the barrier between persons are morally acceptable, in the epistemic case there is no wiggle room: the epistemic separateness of propositions is nonnegotiable. When it comes to the evaluation of individual beliefs, it is never epistemically defensible to sacrifice the furtherance of our epistemic aims with regard to one proposition in order to benefit our epistemic aims with regard to other propositions (even if we grant to the teleologist that there are such things as distinctively epistemic aims and that all epistemic appraisal should be explicated in terms of how well the objects of appraisal conduce toward the furtherance of those aims).⁴¹

39. To secure these results, we may need to increase the number of true beliefs that I would acquire were I to receive the grant, or increase the frequency in the actual world of situations like the one I face (or both). Doing so makes the example more fanciful, but does not change its basic structure.

40. More precisely, I should speak here of "the epistemic separateness of propositions-at-a-time" since it is also epistemically irrelevant whether or not a belief in $\langle p \rangle$ at a given time conduces toward the promotion of true belief and the avoidance of false belief with regard to *that same proposition* at later times. To avoid cumbersome locutions, I will often say "proposition" in what follows when I really mean "proposition-at-a-time."

41. Notice that my objection here applies just as well to a version of veritism according

Two ways of responding to this argument immediately spring to mind. The first response is to hold on to veritism as a theory of final epistemic value but to restrict the teleologist's theory of overall epistemic value so as not to include *causal means* among the ways in which an item in a given evaluative focal point can conduce toward or promote items of final epistemic value. For example, maybe only *constitutive means* count as ways of conducing toward the furtherance of our veritistic aims of believing truths and not believing falsehoods. This restriction would, of course, immediately block any untoward results for the teleologist in the case of the grant-seeking scientist since in that example the scientist's belief in the existence of God is merely a causal means to the acquisition of various true beliefs and the avoidance of various false beliefs sometime in the future. But the question to ask is: *why* should we restrict what counts as conducing toward or promoting value in this way? *What* is the motivation, for the veritistic epistemic teleologist, in making this move (other than to avoid a potential problem)? On the epistemic teleologist's basic picture, certain items (such as true beliefs) have nonderivative epistemic value, and certain other items have derivative epistemic value insofar as they conduce toward the items of that first sort. What is the rationale for restricting this conducing relation so that it doesn't cover causal means? Nothing in the epistemic teleologist's basic picture suggests an answer to this question.⁴²

to which, rather than *all* true beliefs having epistemic value as ends, it is only true beliefs in propositions of interest or importance to us that have epistemic value as ends. A number of authors, including Alston (2005, 32) and Goldman (1999, 88–89; 2001, 38–39), restrict veritism in this way to avoid the worry that trivial truths—such as the phone number of the 323rd person listed in a now-defunct phonebook for Wichita, Kansas—do not seem to be the sorts of things that, from a purely epistemic point of view, we should seek to believe for their own sake. However, conflicts can occur between our veritistic aims with regard to different propositions even when all of the propositions in question are interesting or important, as the example of the grant-seeking scientist shows, so restricting veritism in this way does not help avoid the charge that veritistic teleological theories ignore the separateness of propositions.

42. Moreover, it is worth pointing out that process reliabilism—which, as we saw, is perhaps the most prominent example of a teleological epistemic theory—*does* allow causal means to be a way of conducing toward or promoting final epistemic value. Most forms of process reliabilism restrict these causal means to cases of *proximate causation*, in which a belief-forming process *directly* causes the formation of a true or false belief without the help of any causal intermediaries, but this is still to allow that causation can be a way of promoting epistemic value. I discuss the restriction to proximate causal means below.

A second way of responding to the argument I have presented is to take a hint from Foley. You will recall that Foley (1993, 19) specifies the content of our most basic epistemic goal as “*now* believing those propositions that are true and *now* not believing those propositions that are false” (emphasis mine). So maybe what the teleologist should do is reformulate his or her theory of final epistemic value so that the goals in question—and, correspondingly, the items of value—are time indexed in this manner. Let us call this new version of veritism “time-relative veritism.” (Thus the original version of veritism is more accurately described as “time-neutral veritism.”) Formulated in terms of objects/ends, time-relative veritism holds that (i) for every currently true proposition $\langle p \rangle$, a state of affairs in which the agent *now* believes that p *now* has epistemic value as an end; (ii) for every currently false proposition $\langle q \rangle$, a state of affairs in which the agent *now* believes that q *now* has epistemic disvalue as an end; and (iii) nothing else *now* has epistemic value or disvalue as an end. Formulated in terms of goals/aims, time-relative veritism holds that, at a given time, our only positive ultimate epistemic aims are, for every proposition, to *now* believe that proposition if it is *now* true, and our only negative ultimate epistemic aims are, for every proposition, to *now* not believe that proposition if it is *now* false. Once veritism has been reformulated in this way, cases like the grant-seeking scientist are no longer an obvious embarrassment for the veritistic teleologist: that the scientist will arrive at many true beliefs and have fewer false beliefs *in the future* if he or she believes that God exists is, from the perspective of time-relative veritism, epistemically irrelevant. All that matters epistemically, according to the time-relative veritist, is whether or not the scientist’s belief that God exists promotes true belief and the avoidance of false belief *at the time at which it is held*.⁴³

This move on the veritistic teleologist’s part is structurally identical to a similar move employed by many ethical teleologists who seek to incorporate, within their teleological theory, the verdict that it is impermissible to murder someone in order to prevent five murders from being committed. At first it might seem puzzling how a teleological ethical theory could yield such a result. If one murder is bad, wouldn’t five

43. Though this raises the question: is the time-relative veritist who does not also restrict the conducing relation to preclude causal means committed to the result that, even if the scientist’s belief in the existence of God is initially epistemically unjustified, that belief *becomes* justified once the scientist comes to hold the other beliefs that he or she forms as a result of the grant?

murders be worse, and thus isn't an action that involves committing one murder in order to prevent five murders better overall—and hence more choiceworthy—than an action that involves abstaining from that one murder, thereby allowing the five to occur? (For expository purposes, I assume here that the teleological theory in question is a version of act-consequentialism; the point generalizes.) The first move that the teleologist can make to evade this result is to include both *agent-relative* and *agent-neutral values* in his or her theory of final value and then to insist that murders have an especially strong agent-relative disvalue: from my perspective as an agent confronted with a decision about what to do, the state of affairs in which *I* murder someone has, according to this sort of teleologist, much more disvalue than the state of affairs in which a murder occurs. If this move is defensible,⁴⁴ then it allows the teleologist to handle cases in which I can choose to murder someone in order to stop five other murders from being committed by a person (or group of people) other than myself since the murder I would commit has much more agent-relative plus agent-neutral disvalue than the five other murders I would prevent. However, the move to agent-relative value does not handle cases in which I can choose to murder someone in order to stop *myself* from murdering five other people *sometime in the future* (Kamm 1993, 95; 1996, 242). This kind of case leads teleologists who want to countenance bans on such trade-offs to make a second move, namely, to allow both *time-relative* and *time-neutral values* in their theory of final value and then to insist that murders have an especially strong time-relative disvalue: from my perspective *now* as an agent, the state of affairs in which I *now* murder someone has much more disvalue than a state of affairs in which I murder someone *at a different time*. Or so, at least, some ethical teleologists argue.

Thus the transition from time-neutral to time-relative veritism is an instance of a familiar ploy in the teleologist's bag of tricks.⁴⁵ Again, though, we should ask: what is *the motivation* in making this move? Why is it that current true beliefs have epistemic value as ends, but—mysteriously—future true beliefs do not? Like the move of restricting the conducting relation so as to preclude causal means, this move of time-relativizing our

44. For some doubts, see Ridge 2009, 425–26, following on Kamm 1993.

45. Moreover, the other transition mentioned earlier, namely, from an agent-neutral to an agent-relative conception of value, is one that is already implicitly made by most veritists. Most veritists take epistemic value to be agent-relative: what matters to the epistemic assessment of my belief-forming practices, according to them, is whether those practices promote *my* holding true beliefs and *my* avoiding false beliefs, not whether they promote *anyone's* holding true beliefs and *anyone's* avoiding false beliefs.

ultimate epistemic ends has the feel of an ad hoc maneuver invoked purely to get around a potential objection. Moreover, there is an additional problem with the transition to a time-relative conception of epistemic value: deliberation takes time. When I deliberate about whether to believe that *p*, my aim seems to be to *eventually* come to the right belief on the matter, even if I realize that—because the question is particularly difficult, or my evidence frustratingly equivocal, or my mind currently addled—it might take me quite a while to reach a final verdict. But from the perspective of time-relative veritism, this attitude is misguided: if my ultimate epistemic goal vis-à-vis the proposition $\langle p \rangle$ is to now believe it if it is true and to now not believe it if it is false, then it seems that the goal of my deliberations should be to settle the matter *now*, not (say) ten minutes from now after I have brought my ponderings to a close. But this strikes me as an extremely odd, perhaps even incoherent, attitude to take toward deliberation, which of necessity is a continuous process that takes place over time across a series of distinct *nows*.

I have just mentioned two strategies for replying to my objection that veritistic teleological theories illicitly ignore the separateness of propositions: the first involves excluding causal means from being a way in which epistemic value can be promoted, the second involves moving from a time-neutral to a time-relative conception of our ultimate epistemic aims. In both cases I have expressed some reservations about these responses: neither is well motivated, and the second is in tension with the fact that theoretical deliberation takes place over time. But those are not the most pressing problem with these two responses. No, the most pressing problem is that although these responses allow the teleologist to avoid *some* cases involving illicit cross-propositional trade-offs, they do not allow him or her to avoid *all* such cases.

To see why, consider this. Although theoretical reasoning and deliberation often seems to be concerned with a single proposition—“Is the library open today?” “Will it rain this afternoon?” “Did she do that on purpose?”—as reasoners we rarely reach a verdict about a single proposition at a time, even if only one proposition is the main focus of our attention. Similarly, when forming beliefs in ways other than through explicit reasoning, we rarely—if ever—acquire beliefs one at a time; rather, we tend to acquire batteries of interconnected beliefs about a given subject matter. But if one of these batteries of beliefs contains an inevitable false belief but also many true beliefs, then accepting that battery of beliefs could be a *constitutive means* of sacrificing our veritistic aims with regard to one proposition in order to *at the same time* further our

veritistic aims with regard to a large number of other propositions. So even veritistic epistemic teleologists who restrict themselves to constitutive means or to a time-relative theory of final epistemic value must concede that there can be cases in which it is epistemically appropriate, by their lights, to sacrifice one proposition for the greater (epistemic) good.

Here is one case of that sort. Suppose I am trying to figure out a paradox such as the liar paradox, or the lottery paradox, or one of the paradoxes that plague deontic logic. Formally, a paradox is a set of propositions that are individually plausible but jointly inconsistent. In this case, let us suppose that the paradox in question consists of three distinct propositions. Let us also suppose that all three propositions are equally compelling, with none an obvious candidate for the source of inconsistency. Moreover, to make the case a bit cleaner, let us stipulate that I have excellent evidence that only one of the three propositions is false: maybe my friend, who knows the solution to the paradox but wants me to figure it out for myself, has told me that only one of them is false, or maybe the overwhelming plausibility of each of the propositions is sufficient to make it extremely unlikely that more than one of them is untrue. Finally, let us fill in the story as follows: before I realized them to be incompatible, I believed all three propositions. Once their incompatibility is pointed out to me, I spend some time trying to figure out which one of them is false. After a few frustrating minutes, I think to myself, "Forget it, this philosophy stuff is too hard," and go back to believing all three propositions. The question is: am I epistemically justified, or reasonable, or rational, in doing so?

I think it is clear that I am not. What I should do is suspend judgment on the three propositions until I am able to single out one (or two) of them as the least plausible of the bunch. Then I should either suspend judgment on those propositions or outright disbelieve them, depending on the strength of my case against them.⁴⁶ The problem, though, is that it

46. Some readers may think that, by insisting that I am not justified in going back to believing all three propositions, I commit myself to an implausible response to the paradox of the preface. But no such thing follows, for there are crucial disanalogies between the case at hand and the scenario envisaged in the paradox of the preface. In the preface paradox, an author of a book has, for each claim made in his or her book, some inductive evidence that the claim is false, which together give him or her excellent inductive evidence that one of the book's claims is false. In the case at hand, I have noninductive evidence that one of three propositions is false. So to make my belief that one of these three propositions is false truly analogous to our author's prefatory remark, we'd need to consider a variant of the preface paradox in which the author's book consists of one

is difficult to see how veritism—even a time-relative version of veritism that bans causal means—can yield this result. In going back to believing all three propositions, I take on a belief in a falsehood but also thereby come to believe two truths. From the veritist's perspective, shouldn't my failing in my epistemic aims with regard to the one false proposition be outweighed by my success in my epistemic aims with regard to the other two propositions?⁴⁷

The lesson of this example is that even a time-relative veritist who permits only noncausal ways of promoting value must sanction implausible trade-offs between our veritistic aims with regard to different propositions.⁴⁸ And another sort of example can illustrate the same lesson in an even more convincing manner if we allow ourselves to make a certain (somewhat contentious) assumption in the philosophy of mind. The assumption in question concerns the nature of beliefs about one's own beliefs.⁴⁹ Some authors—incorrectly, to my mind—think that higher-order beliefs and the lower-order beliefs featured in those higher-order beliefs are distinct existences, so that one's belief in $\langle p \rangle$ can stand

continuous philosophical argument that eventually leads to a contradiction. But retaining belief in that book's premises and writing in its preface "At least one of the premises of this book's central argument is false since that argument leads to a contradiction" is not paradoxical; it is foolhardy.

Also, it is worth pointing out that those who think that, in the case at hand, I am justified in retaining belief in three propositions *that I recognize* to be inconsistent face the following dilemma: either they must give up on the idea that we can always acquire new justified beliefs by deductively reasoning from our old justified beliefs (what is known as "multipremise closure for justified belief"), or they must accept that after I go back to believing all three propositions, I can become justified in believing any arbitrary proposition (since an inconsistent set of propositions entails everything). I find neither option particularly palatable.

47. This will not follow if a veritist holds that the disvalue of a false belief is more than twice the value of a true belief. But then we can increase the number of propositions making up our paradox to get the desired result. To hold that a false belief is *infinitely* more disvaluable than a true belief is to give up on the Jamesian insight that motivated a move to two sorts of veritistic aims.

48. I think there is a clear sense in which, in the example just given, my believing the false proposition (whichever it is) is a noncausal means to my believing the two true propositions. But what sort of a noncausal means is this, exactly? Here is one proposal: my believing the false proposition is an upward constitutive means to my believing all three propositions, and my believing all three propositions is a downward constitutive means to my believing each of the two true propositions.

49. More precisely, it concerns the nature of those beliefs about one's own beliefs that are formed at least partially on the basis of introspection. I suppress this qualification in what follows.

in causal relations to one's belief in $\langle I \text{ believe } \langle p \rangle \rangle$.⁵⁰ But on another account—which I favor—higher-order beliefs are partially constituted by the lower-order beliefs that they are about, and hence the relation that one's belief in $\langle p \rangle$ stands to one's belief in $\langle I \text{ believe } \langle p \rangle \rangle$ is not the *causal* means-end relation but rather the *constitutive* means-end relation.⁵¹ Let us, for the time being, assume that this latter account is correct; after presenting my problem case for time-relative, noncausal veritists with such an assumption in place, I will show how it is possible to construct a similar problem case even if we reject constitutive accounts of higher-order belief.

On to the example. Suppose that, due to a quirk in my psychology, I form beliefs in a very self-conscious manner. Whenever I come to believe $\langle p \rangle$, where $\langle p \rangle$ is some claim about the external world around me, I also—at the same time—form a series of second-order beliefs about the nature of that first-order belief and the way in which it was formed: $\langle I \text{ believe } \langle p \rangle \rangle$, $\langle I \text{ formed my belief in } \langle p \rangle \text{ on such-and-such day} \rangle$, $\langle I \text{ formed my belief in } \langle p \rangle \text{ using such-and-such method} \rangle$, and so on. Moreover, for each of these second-order beliefs, I—at the same time—form a series of third-order beliefs about that second-order belief: $\langle I \text{ believe } \langle I \text{ believe } \langle p \rangle \rangle \rangle$, $\langle I \text{ formed my belief in } \langle I \text{ believe } \langle p \rangle \rangle \text{ on such-and-such day} \rangle$, $\langle I \text{ formed my belief in } \langle I \text{ believe } \langle p \rangle \rangle \text{ using such-and-such method} \rangle$, and so on. And, for each of these third-order beliefs, I—at the same time—form a series of fourth-order beliefs about that third-order belief: $\langle I \text{ believe } \langle I \text{ believe } \langle I \text{ believe } \langle p \rangle \rangle \rangle \rangle$, and so on. (Maybe this hierarchy continues a few more orders, or maybe the fourth order is where it peters out; for the purposes of my example, it doesn't matter.) Now here is the crucial thing: although I am not particularly adept at forming first-order beliefs about the external world, I am very, very good at introspection and almost always have accurate higher-order beliefs about the nature of my lower-order beliefs and the ways in which they are formed.

Now suppose that, one day, I come to believe the proposition $\langle \text{She loves me} \rangle$ by picking petals from a daisy while reciting, "She love me, she loves me not." In forming this belief, I also come to believe $\langle I \text{ believe } \langle \text{She loves me} \rangle \rangle$, $\langle I \text{ came to believe } \langle \text{She loves me} \rangle \text{ by picking petals from a daisy} \rangle$, $\langle I \text{ came to believe } \langle \text{She loves me} \rangle \text{ on a gloriously sunny Saturday afternoon} \rangle$, and so on, up the hierarchy.

50. The locus classicus for such a view is Armstrong 1968.

51. The locus classicus for such a view is Shoemaker 1994.

Since my powers of introspection are so profound, all of these higher-order beliefs are true. Now clearly my first-order belief is unjustified; if basing amatory beliefs on the outcome of petal picking is not an improper way of forming beliefs, then what is? But, given our assumption about the nature of higher-order beliefs, that first-order belief ended up being a constitutive means to forming, at the same time, a vast number of true beliefs: if n is the number of higher-order beliefs that I, in general, form about each belief of one-lower order, then my first-order belief allowed me to form $n + n^2 + n^3$ true higher-order beliefs. So even a time-relative veritist who forswears causal means must deem my belief about whether she loves me to be a sacrifice for the greater epistemic good, and hence justified.⁵² But a better candidate for an unjustified belief is difficult to come by.

The previous example worked as it did only because we were assuming a constitutive account of the relation between higher- and lower-order beliefs. What, though, if we relax that assumption? Well, then my belief in the proposition <She love me> will not be a constitutive means to my believing <I believe <She loves me>>. However, what we are concerned with here is not just whether my first-order belief is a constitutive means to a state of affairs in which I possess a certain higher-order belief, but more importantly whether my first-order belief is a constitutive means to a state of affairs in which I possess a certain *true* higher-order belief. And notice that my belief in the proposition <She love me> is, in fact, a constitutive means to my truly believing <I believe <She loves me>>: a state of affairs in which I truly believe <I believe <She loves me>> is partially constituted by a state of affairs in which I believe <She loves me>. Thus, regardless of what assumptions we make about the nature of higher-order beliefs, my first-order belief in <She loves me> will be a constitutive means to my having all n of my true second-order beliefs. This same strategy, though, will not work for my

52. This result follows if, as most epistemic teleologists do, we are working with a satisficing deontic theory. However, Anil Gupta has pointed out to me that the result may not follow if we are working with a maximizing deontic theory since even if my belief in <She loves me> promotes a great deal of overall epistemic value, a belief instead in <She doesn't love me> might promote more overall epistemic value, by better promoting epistemic value at the first-order level and promoting as much epistemic value at higher-order levels. But even if that is so, we can avoid this problem by tweaking our example a bit, so that it is built into the case that—for some reason—I would only form a hierarchy of higher-order beliefs if I believe <She loves me> and not if I believe its negation.

beliefs at the third order and higher: even if a state of affairs in which I truly believe $\langle$ I believe $\langle$ I believe $\langle$ She loves me $\rangle\rangle\rangle$ obtains, that state of affairs is not constituted in part by my believing $\langle$ She loves me $\rangle$ if we renounce a constitutive account of higher-order beliefs. However, to get around this setback, we need only vary our example a bit: simply make all of my higher-order beliefs at the third level and beyond be about what I *truly believe*, rather than about what I *believe*. So let us suppose that, at the third level, I believe $\langle$ I truly believe $\langle$ I believe $\langle$ She loves me $\rangle\rangle\rangle$, $\langle$ I came to truly believe $\langle$ I believe $\langle$ She loves me $\rangle\rangle$ through introspection $\rangle$, and so forth; that, at the fourth level, I believe $\langle$ I truly believe $\langle$ I truly believe $\langle$ I believe $\langle$ She loves me $\rangle\rangle\rangle\rangle$, and so forth; and so on. Now the state of affairs in which I truly have any one of these beliefs will be partially constituted by a state of affairs in which I believe $\langle$ She loves me $\rangle$. The result: we have trouble even for those time-relative, noncausal veritists who are not fans of constitutive accounts of higher-order beliefs.

Therefore moving to a time-relative conception of our veritistic aims or excluding causal means does not avoid the basic problem for veritistic approaches to epistemic teleology. Even then, there can exist conflicts and competition between the veritist's epistemic goals with regard to different propositions. And where such conflict or competition exists, it seems that the teleologist must favor the many over the one, and thus sanction cross-propositional trade-offs. Even time-relative, non-causal veritism does not take seriously the separateness of propositions.

What about other, more complicated ways of tinkering with the details of one's teleological theory to avoid undesirable trade-offs? At this point, it is worth considering one particularly salient example in that vein, namely, the process reliabilist's way of formulating a veritistic teleological theory. Two features of process reliabilism are pertinent here. First, process reliabilism contains an indirect deontic theory in which belief-forming processes are *directly* assessed in terms of how well they promote epistemic value, and individual beliefs are *indirectly* assessed in terms of their connection to belief-forming processes that pass a certain muster. Second, most forms of process reliabilism include a theory of overall value in which the conducing relation is restricted so that only *proximate causal means* are a way of promoting epistemic value: when assessing the overall epistemic value of a belief-forming process, only beliefs formed *immediately* by a given instantiation of that process, without the aid of any causal intermediaries, are taken to be relevant. This rules out both *nonproximate causal means* (such as those featured in the grant-seeking

scientist case) and *noncausal means* (such as those featured in the paradox case and in the case of the incredibly accurate introspector) as ways of promoting epistemic value. For this reason, none of the trade-off cases I have discussed so far pose a problem for the process reliabilist.

Nevertheless, much the same dialectic as before plays out. First, the process reliabilist's restriction of the conducting relation is under-motivated, given his or her general teleological commitments. Why, exactly, are nonproximate causal means banned from being a way of promoting epistemic value, whereas proximate causal means are not? Such a fine distinction has an air of arbitrariness to it.⁵³ But, more importantly, the features of process reliabilism that allow it to avoid *some* cases involving objectionable cross-propositional trade-offs do not allow it to avoid *all* such cases.

For example, consider the following belief-forming process. Suppose that whenever I consider whether a given natural number n is prime, I form the following belief: $\langle n$ is not prime $\rangle$. This process tends to yield a ratio of true to false beliefs that approaches 1.⁵⁴ Thus, from the process reliabilist's perspective, my belief-forming process is almost perfectly reliable, and the false beliefs it occasionally yields are sacrifices for the (much) greater epistemic good. But when I form a belief in $\langle 7$ is not prime $\rangle$ on the basis of this process, I take it to be a datum that my belief is not justified (or reasonable, or rational). The lesson is that even veritists of a process-reliabilist persuasion allow illicit cross-propositional trade-offs between our epistemic goals, although these trade-offs occur at the level of belief-forming processes rather than at the level of individual beliefs.⁵⁵

53. Moreover, it is worth pointing out that, in other contexts, process reliabilists are perfectly willing to allow nonproximate causal means to transmit epistemic value. In particular, the first (and, to my mind, more promising) of Goldman and Olsson's (2009) two solutions to the so-called swamping problem for process reliabilism only works if nonproximate causal means are a way of transmitting epistemic value.

54. It follows from the Prime Number Theorem that the density of primes less than x approaches 0 as x approaches infinity.

55. Two ways of resisting this result suggest themselves. The first is to insist that, when I come to believe that 7 is not prime, the relevant process by which my belief is formed is more narrow than the process of forming a belief in $\langle n$ is not prime $\rangle$ for any natural number n . In a sense, this sort of response is always available to process reliabilists when faced with a putative counterexample, precisely because of another problem that bedevils their view, namely, the generality problem (see Feldman 1985; Conee and Feldman 2004 [1998]; and Feldman and Conee 2002). But such a maneuver only deflects this particular example and not the more basic problem: for any given way of specifying the

I can imagine one final, last-ditch effort on the part of the veritist.⁵⁶ Perhaps, in addition to moving to a time-relative conception of epistemic value, we should also move to a *proposition-relative* conception of epistemic value. The idea is that a state of affairs in which I believe true proposition $\langle p \rangle$ has final epistemic value relative to the proposition $\langle p \rangle$ but no epistemic value relative to any other propositions, just as—according to many ethical teleologists who are fans of time-relative value—a state of affairs in which I *now* murder someone has final disvalue relative to the moment of time picked out by the indexical “now” and none relative to any other time. Would revising veritism in this way allow the epistemic teleologist to avoid having to countenance cross-propositional trade-offs?

No, it would not. Relativizing the values that make up a teleologist’s theory of final value in a given way does not, by itself, allow the teleologist to avoid having to countenance certain trade-offs. Rather, relativizing values only helps when, as a result of that relativization, certain cases of conflict or competition among values no longer arise. But in the case of this move to a proposition-relative conception of epistemic value, the relativization in question does not decrease the number of cases of conflicts or competition between items of final epistemic value or disvalue since one can come to believe that p at the same time as one

narrowness or broadness of the belief-forming process that, according to process reliabilism, is relevant to the epistemic status of a given belief, it will be possible to construct problem cases with exactly the same structure as the one provided here.

The second way of resisting this case is to revise process reliabilism’s deontic theory so that what matters is not just whether a belief is formed on the basis of a reliable process, but also whether there is some other belief-forming process available to the subject that, had it been used, would have resulted in that belief’s not being held. See Goldman 1992 [1979], 123. More precisely, the suggestion is that process reliabilism’s deontic subtheory for the deontic focal point {beliefs} and the deontic property *being justified* be revised as follows:

S’s belief that p at time t is justified iff (i) the belief-forming process that caused S to believe that p at t is reliable, and (ii) there is no reliable belief-forming process available to S that, if it had been used in addition to the process actually used, would have resulted in S’s not believing that p at t .

However, this revision helps with the case at hand only if it is true of me that there is another reliable process available to me that, if used, would result in my not believing $\langle 7 \text{ is not prime} \rangle$. And we can simply stipulate that this is not true in our case. Moreover, revising process reliabilism in the proposed manner leads to well-known problems: see Bonjour 1985, 47–49; and Feldman 1985, 165–67.

56. I thank Michael Smith for suggesting to me a version of the objection to follow.

comes to, or continues to, believe that q (where $\langle p \rangle$ and $\langle q \rangle$ are distinct propositions). So even proposition-relative veritism must weigh the furtherance or frustration of our epistemic aims with regard to one proposition against the furtherance or frustration of our epistemic aims with regard to other propositions. A proposition-relative theory of final epistemic value is not a way of imposing epistemic separateness between propositions.

An analogy will help bring out the point. Suppose an ethical teleologist holds a theory according to which, in certain cases, we are morally obligated to torture one person to death in order to prevent a very large number of people from each getting a paper cut. Could the teleologist avoid this (let us assume) embarrassing result by moving to a *harm-relative* theory of final value according to which we have one ultimate negative goal of, with regard to the harm of torturing someone to death, not bringing about *a state of affairs in which someone is tortured to death* and another ultimate negative goal of, with regard to the harm of receiving a paper cut, not bringing about *a state of affairs in which someone receives a paper cut*? No, he or she could not. Even if these goals somehow have as part of their content that they pertain only to a given harm, the two goals can be satisfied or frustrated at the same time, so altering their content in this way does not cause them to cease to be in conflict with one another in a case in which one must choose between torturing one person to death and allowing a vast number of people to each receive a paper cut.⁵⁷ So, too, in the case of a move to epistemic goals that have, as part of their content, the fact that they pertain only to a given proposition: these goals can be fulfilled or frustrated at the same time, so altering their content in this way does not help the teleologist avoid having to say something about cases in which the promotion of a veritistic goal with regard to one proposition can be sacrificed in order to benefit the promotion of a vast number of veritistic goals with regard to other propositions.

This completes my case against veritistic teleological theories. To recap: Firth-style cases such as the grant-seeking scientist illustrate how the crudest forms of veritistic teleology are forced to sanction sacrificing our epistemic aims with regard to one proposition in order to further our epistemic aims with regard to many other propositions, thereby ignoring the epistemic separateness of propositions. We considered various ways of

57. A similar point holds if, rather than considering a *harm-relative* theory of final value, we instead consider a *victim-relative* theory of final value. I have chosen the former because it more vividly makes my point.

complicating a veritistic teleological theory, whether by relativizing its theory of final value or by restricting the conducting relation in its theory of overall value, but none of these maneuvers managed to avoid the central problem: it was still possible to construct cases in which objectionable cross-propositional trade-offs determine the epistemic status of an individual belief. My conclusion is that any vaguely plausible variety of veritism must, in some cases, disregard the distinctness of propositions.⁵⁸

What, though, about nonveritistic teleological theories? Are they tenable? The argument I have just offered generalizes to almost all teleological epistemic theories in a fairly natural way. First, consider teleological epistemic theories that take true beliefs to have epistemic value as ends and false beliefs to have epistemic disvalue as ends, but also deem other items—such as coherent systems of belief, or understanding, or wisdom—to have epistemic value as ends. (This is, without a doubt, the most common sort of nonveritistic teleological theory found in the literature.) Since the ultimate epistemic goals countenanced by these theories include all of the ultimate epistemic goals countenanced by the veritist, my earlier argument works against these theories as well: in cases in which we can sacrifice our veritistic goal of believing the truth with regard to one proposition in order to greatly further our veritistic goals with regard to a large number of other propositions, these theories will flout the separateness of propositions by deeming such sacrifices to be epistemically appropriate.⁵⁹

58. Why the qualification “vaguely plausible”? Answer: because there are some highly implausible veritistic theories that avoid my argument. One involves restricting the conducting relation so that instantiation is the only way of promoting epistemic value. Another involves constructing a deontic theory only for the focal point {complete systems of belief} and not for the focal point {beliefs}. Against the former: then the theory’s teleological structure is doing no work (and then it is forced to equate justified beliefs with true beliefs, but presumably not all true beliefs are justified). Against the latter: this is to give up on answering the question “What should I believe?” (that is, to give up on epistemically assessing individual beliefs).

59. Strictly speaking, it might turn out that, in some of these cases, an agent can only sacrifice the truth-in-belief goal with regard to the one proposition at the cost of thwarting some of his or her other, nonveritistic goals. (For example, maybe falsely believing the proposition to be sacrificed would lower the coherence of the agent’s entire system of belief, but truly believing the other propositions would not raise—or would not raise as much—the coherence of his or her belief-system.) But even if this is so in some cases, it would be mind-bogglingly convenient if in *every* case of cross-propositional conflict between one’s veritistic aims, one’s nonveritistic aims exactly compensate for those conflicts. Compare: this would be like an act-utilitarian who defends a pluralist conception of well-being saying that his or her theory does not ignore the separateness of persons

This leaves one last possibility: teleological theories that do not accept that true beliefs have final epistemic value or do not accept that false beliefs have final epistemic disvalue. For example, maybe according to such a theory it is only knowledge, or only coherent systems of belief, or only understanding, that is to be promoted as an epistemic end for its own sake. But, since in all of these theories the bearers of fundamental value are cognitive states that one can attain with respect to one proposition (or set of propositions, or domain of inquiry) while not attaining with respect to various other propositions (or sets of propositions, or domains of inquiry), we can run a version of my argument against such teleological theories as well.

For example, suppose our teleologist's theory of final value holds that the only items of epistemic value as ends are coherent systems of belief (where the degree of value is proportional to the degree of coherence among the beliefs), and the only items of epistemic disvalue as ends are systems of belief that lack coherence (where the degree of disvalue is proportional to the degree to which the beliefs fail to cohere with one another). The argument against this view will be almost identical to my argument against veritistic theories. The opening move will be the same: the grant-seeking scientist, for example, sacrifices coherence with regard to his or her God-related beliefs in order to greatly increase the coherence of his or her scientific beliefs. In order to avoid having to say that the scientist's belief in God's existence is epistemically justified, our teleologist might restrict the promoting value relation to rule out causal means or might move to a time-relative conception of the epistemic goal of attaining coherence (and avoiding incoherence) with regard to systems of belief. But then, just as before, there will be cases in which the coherence of one small patch of one's web of belief can be sacrificed via constitutive means in order to, at the same time, greatly increase the coherence of a large number of other patches of one's web of belief. For example, the incredibly accurate introspector slightly sacrifices the coherence of his or her romantic beliefs in order to acquire a beautifully interwoven, highly coherent fabric of beliefs about his or her own belief-forming practices on a given occasion. Thus even after the teleologist reformulates his or her theory, the basic problem remains. And so on.

because in every case in which we can choose to trade one person's pain for several other people's pleasure, there are compensating losses of nonhedonic well-being for those other people (such as a loss of dignity).

Because a gain in coherence in subsystem B_1 of one's belief-system can come at the cost of a loss in coherence of subsystem B_2 of one's belief-system, where these two subsystem do not contain any propositions in common, even teleological theories whose theory of final value is formulated purely in terms of coherence must sanction trade-offs that cross the boundaries between propositions.

6. Conclusion: Beyond Truth-Conduciveness

I have just argued that most teleological epistemic theories ignore the separateness of propositions and for that reason should be rejected. However, the argument I have provided so far does not—strictly speaking—apply to *every possible* teleological account of epistemic notions: there are moves and countermoves, ploys and counterploys, that I have not had the space to consider. But as the teleologist continues to pile on the qualifications and cobble together yet more codicils in a effort to explain away each new case involving objectionable trade-offs, I am left wanting to respond: “Sure, you can defend that qualification-heavy, codicil-ridden theory if you want to; but why would you want to? Why start with a framework whose hallmark is its propensity for trade-offs and then work so hard to erase that feature? Why not just start somewhere else?”

So let us end by considering—all too briefly—two final issues: (i) What are the alternatives to a teleological approach to normative epistemology? and (ii) What is the theoretical pay-off once we renounce the teleological approach?

First, then: what's the alternative? As it turns out, a number of nonteleological alternatives already exist in the literature, provided that we divest them of their oft-accompanying but nonessential teleological accoutrements. I mention here just two examples. First, one might defend evidentialism about epistemic justification together with a non-teleological theory of evidence. By “evidentialism,” I mean (roughly) the view that belief in proposition $\langle p \rangle$ is justified for a given subject at a given time if and only if the subject's evidence at that time for and against $\langle p \rangle$ on balance supports $\langle p \rangle$.⁶⁰ When formulated in this bare-bones

60. This is close to what Earl Conee and Richard Feldman mean by “evidentialism” in a well-known series of articles defending such a view; see the papers collected in Conee and Feldman 2004, as well as Conee and Feldman 2008. Conee and Feldman occasionally appeal to teleological notions when defending their brand of evidentialism: for example, Feldman (2004 [2000], 181–86) tentatively proposes that it maximizes epistemic value to follow one's evidence, and Conee (2004 [1992], 248–51) invokes the epistemic goal of

manner, evidentialism doesn't have much content since "evidence" is itself a normative notion, and there are many, many different theories of evidence.⁶¹ But if we plug into our evidentialist formula a nonteleological theory of evidence, then the resulting theory will be nonteleological.

My second example of a nonteleological approach to epistemic normativity can be found in the work of John Broome. For the past decade, Broome has delineated and defended a series of rational requirements that each take the form of a wide-scope principle exhorting us to have, or not to have, a certain combination of attitudes (see Broome 1999, 2001, 2005, 2007, 2008, and forthcoming). Most of these rational requirements are mixed practical/theoretical requirements, featuring both practical attitudes such as intention and theoretical attitudes such as belief; for example:

- (R₁) Rationality requires that [if (one intends that *e*, and one believes that it will only be the case that *e* if it is the case that *m*, and one believes that it will only be the case that *m* if one intends that *m*), then one intends that *m*].

However, some of Broome's rational requirements are wholly practical, such as

- (R₂) Rationality requires that it not be the case that [one intends to ϕ , and one intends not to ϕ],

and others are wholly theoretical, such as

- (R₃) Rationality requires that it not be the case that [one believes that *p*, and one believes that not-*p*].

The crucial point for my purposes is that Broome doesn't think that rationality requires us to have or not have these combinations of attitudes because doing so conduces toward something of value; rather, he just thinks that that's what rationality requires. So Broome's wholly theoretic-

maximizing one's total stock of knowledge when discussing a particular account of the connection between epistemic justification and truth. However, both of these arguments are problematic, for precisely the sorts of reasons detailed in sec. 5 above, and I think Conee and Feldman would be better served by jettisoning any talk of epistemic value or epistemic goals from their overall position.

61. For a sampling of some of the possibilities here, see Kelly 2008a and 2008b.

cal rational requirements would count as a nonteleological approach to epistemic rationality.⁶²

Since the two sorts of epistemic theories I have just mentioned are nonteleological in structure, it is tempting to refer to them as examples of what might be called “epistemic deontology.” However, that temptation should be resisted, for two reasons. First, the name “epistemic deontology” has already been appropriated by epistemologists for another purpose. In his famous essay of the same name, William Alston (1989 [1988a], 115–18) uses the phrase “the deontological conception of epistemic justification” to refer to the view that epistemic justification can be cashed out in terms of epistemic permission, obligation, duty, responsibility, blameworthiness, or praiseworthiness.⁶³ Now Alston’s choice of terminology here is unfortunate and has led to much confusion over the years: a more faithful term for the position he was trying to capture would be “the deontic conception of epistemic justification” (‘deontic’ means “pertaining to duty or obligation,” whereas ‘deontological’ means “pertaining to *the study of* duty or obligation”).⁶⁴ Indeed, one can be an epistemic nonteleologist (in my sense) without endorsing the deontological conception of epistemic justification (in Alston’s sense),⁶⁵ and one can be an epistemic teleologist (in my sense) while embracing the deontological conception of epistemic justification (in Alston’s sense).⁶⁶ But at this point Alston’s terminology is too well entrenched to resist. The second reason not to refer to nonteleological epistemic theories as examples of “epistemic deontology” is that, in my opinion, ‘deontology’

62. Assuming, that is, that epistemic rationality just is theoretical rationality. For a contrary view, see Kelly 2003, 634–37.

63. See also Alston 1989 [1985], 84–86.

64. Moreover, it is simply a mistake on Alston’s part to assimilate *hypological notions* such as responsibility, blameworthiness, and praiseworthiness and *deontic notions* such as permission, obligation, and duty. (The useful term “hypological”, from the Greek *hypologos* for ‘hold accountable or liable’, I owe to Michael J. Zimmerman [2002, 554].) Although it is standard to assume that deontic notions are interdefinable in a fairly straightforward manner (‘S is obligated to ϕ ’ being logically equivalent to ‘S is not permitted to not ϕ ’, and so on), the exact connection—if any—between deontic and hypological notions is a more controversial matter, whether in the case of action or belief. So really we should distinguish *the deontic conception of epistemic justification* from *the hypological conception of epistemic justification*.

65. Perhaps one is a nonteleologist who holds that epistemic justification has to do with a certain sort of fittingness (where this fittingness is not in turn analyzed in terms of a deontic modal such as ‘permitted’ or ‘obligated’).

66. Perhaps one is a teleologist who holds that ‘... is epistemically justified’ and ‘... is permitted by the norms of epistemic rationality’ are synonymous.

is no longer a useful way of designating theories, even in the ethical realm: at this point, there is such a wide array of radically different theories falling under the term ‘deontology’ that the time has come to stop pretending that they form a unified category. Kantianism, Rossianism, contractualism, Kamm-style nonconsequentialism, side-constraint theories, even moral particularism: all of these could, with some justification, be referred to as forms of deontology, yet they have little in common with one another other than the fact that they are not versions of consequentialism.

I have mentioned two varieties of nonteleological epistemic theories, and I have urged that they not be referred to as instances of “epistemic deontology.” But regardless of what we call the nonteleological approach to epistemic normativity, does accepting that approach have any theoretical pay-off in other areas of philosophy? I believe that it does. In particular, I believe that the rejection of epistemic teleology can make a number of far-reaching problems in ethics and epistemology that currently seem intractable a bit more tractable. The teleological approach to epistemic normativity is so pervasive that, in many circles, the predicate ‘... is a source of epistemic justification’ has almost become synonymous with the predicate ‘... is truth-conducive.’ So when people ask, “Are intuitions trustworthy?” or “Do we have good reason to believe the results of reflective equilibrium?” or “Is inference to the best explanation a way of acquiring justified beliefs?” they often take these to be questions about whether the belief sources under discussion tend to result in mostly true beliefs. But showing that one of these belief sources tends to result in mostly true beliefs is extremely difficult, if not impossible, especially if we deem it illegitimately circular to appeal to that very belief source during our proof of its truth-conduciveness. For example, I think there is no hope of showing that *intuitions about distinctly philosophical subject matters* are truth-conducive without appealing, at some point, to other intuitions of that same sort.⁶⁷ In the case of *reflective equilibrium*, it is now widely agreed that if you start with a sufficiently distorted set of initial considered judgments, especially your higher-order considered judgments about the comparative plausibility of your lower-order considered judgments, there is little chance that you will end up with true beliefs when you finally reach

67. As, for example, George Bealer (1996a, 127; 1996b, 11; 1998, 217) does during his highly sophisticated defense of the reliability of philosophical intuitions.

an equilibrium point (if, indeed, you ever do).⁶⁸ And, finally, in the case of *inference to the best explanation*, if we focus just on the explanatory virtue of simplicity, the history of attempts to show that simpler theories are more likely to be true doesn't inspire one with confidence, to put it mildly.⁶⁹ My hope is that moving to a nonteleological conception of epistemic normativity will make it easier to show that it is epistemically appropriate to trust belief sources such as these. But that, of course, is a project for another day.

References

- Adams, Robert Merrihew. 1976. "Motive Utilitarianism." *Journal of Philosophy* 73: 467–81.
- Alston, William P. (1976) 1989. "Self-Warrant: A Neglected Form of Privileged Access." *American Philosophical Quarterly* 13: 257–72. Reprinted in Alston 1989, 286–315. Page references are to the 1989 reprint.
- . 1989 [1985]. "Concepts of Epistemic Justification." *Monist* 68: 57–89. Reprinted in Alston 1989, 81–114. Page references are to the 1989 reprint.
- . 1989 [1986]. "Internalism and Externalism in Philosophy." *Philosophical Topics* 14: 179–221. Reprinted in Alston 1989, 185–226. Page references are to the 1989 reprint.
- . 1989 [1988a]. "The Deontological Conception of Epistemic Justification." *Philosophical Perspectives* 2: 257–99. Reprinted in Alston 1989, 115–52. Page references are to the 1989 reprint.
- . 1989 [1988b]. "An Internalist Externalism." *Synthese* 74: 265–83. Reprinted in Alston 1989, 227–45. Page references are to the 1989 reprint.
- . 1989. *Epistemic Justification: Essays in the Theory of Knowledge*. Ithaca, NY: Cornell University Press.
- . 1991. *Perceiving God: The Epistemology of Religious Experience*. Ithaca, NY: Cornell University Press.
- . 1993. *The Reliability of Sense Perception*. Ithaca, NY: Cornell University Press.
- . 2005. *Beyond "Justification": Dimensions of Epistemic Evaluation*. Ithaca, NY: Cornell University Press.
- Anderson, Elizabeth. 1993. *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.
- Anscombe, G. E. M. 1958. "Modern Moral Philosophy." *Philosophy* 33: 1–19.

68. A point made particularly forcefully by Michael DePaul (1993, 23–56; 1998, 297–98).

69. See Baker 2011 for a helpful overview.

- Armstrong, D. M. 1968. *A Materialist Theory of the Mind*. London: Routledge.
- Baergen, Ralph. 1995. *Contemporary Epistemology*. Fort Worth, TX: Harcourt Brace.
- Baker, Alan. 2011. "Simplicity." In *Stanford Encyclopedia of Philosophy*, summer 2011 ed., ed. Edward N. Zalta. plato.stanford.edu/archives/sum2011/entries/simplicity/.
- Banner, William Augustus. 1968. *Ethics: An Introduction to Moral Philosophy*. New York: Charles Scribner's Sons.
- Bealer, George. 1996a. "A Priori Knowledge and the Scope of Philosophy." *Philosophical Studies* 81: 121–42.
- . 1996b. "On the Possibility of Philosophical Knowledge." *Philosophical Perspectives* 10: 1–34.
- . 1998. "Intuition and the Autonomy of Philosophy." In DePaul and Ramsey 1998, 201–39.
- Bernecker, Sven, and Duncan Pritchard, eds. 2011. *The Routledge Companion to Epistemology*. London: Routledge.
- Blumenfeld, David, and Jean Beer Blumenfeld. 1978. "Can I Know That I Am Not Dreaming?" In *Descartes: Critical and Interpretative Essays*, ed. Michael Hooker, 234–55. Baltimore: Johns Hopkins University Press.
- BonJour, Laurence. 1976. "The Coherence Theory of Empirical Knowledge." *Philosophical Studies* 30: 281–312.
- . 1978. "Can Empirical Knowledge Have a Foundation?" *American Philosophical Quarterly* 15: 1–13.
- . 1980. "Externalist Theories of Empirical Knowledge." *Midwest Studies in Philosophy* 5: 53–73.
- . 1985. *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.
- . 1986. "A Reconsideration of the Problem of Induction." *Philosophical Topics* 14: 93–124.
- . 1998. *In Defense of Pure Reason*. Cambridge: Cambridge University Press.
- . 1999a. "The Dialectic of Foundationalism and Coherentism." In *The Blackwell Guide to Epistemology*, eds. John Greco and Ernest Sosa, 117–42. Malden, MA: Blackwell.
- . 1999b. "Foundationalism and the External World." *Philosophical Perspectives* 13: 229–49.
- . 2002. "Internalism and Externalism." In Moser 2002, 234–63.
- . 2003. "A Version of Internalist Foundationalism." In Laurence BonJour, and Ernest Sosa, *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*, 3–96. Malden, MA: Blackwell.
- . 2010. *Epistemology: Classic Problems and Contemporary Responses*. 2nd ed. Lanham, MD: Rowman and Littlefield.
- Booth, Anthony. 2006. "Can There Be Epistemic Reasons for Action?" *Grazer Philosophische Studien* 73: 133–44.

- . 2008. "Deontology in Ethics and Epistemology." *Metaphilosophy* 39: 530–45.
- Brandt, Richard B., ed. 1961. *Value and Obligation: Systematic Readings in Ethics*. New York: Harcourt, Brace, and World.
- Broad, C. D. 1930. *Five Types of Ethical Theory*. London: Routledge.
- Brogaard, Berit. 2009. "The Trivial Argument for Epistemic Value Pluralism, or, How I Learned to Stop Caring about Truth." In Haddock, Millar, and Pritchard 2009, 284–305.
- Broome, John. 1999. "Normative Requirements." *Ratio* 12: 398–419.
- . 2001. "Normative Practical Reasoning." *Proceedings of the Aristotelian Society Supplementary Volume* 75: 175–93.
- . 2005. "Does Rationality Give Us Reasons?" *Philosophical Issues* 15: 321–37.
- . 2007. "Requirements." In *Hommage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, ed. Toni Rønnow-Rasmussen, Björn Petersson, Jonas Josefsson, and Dan Egonsson. www.fil.lu.se/hommageawlodek/site/papper/BroomeJohn.pdf.
- . 2008. "Is Rationality Normative?" *Disputatio* 11: 153–71.
- . Forthcoming. *Rationality through Reasoning*. Oxford: Oxford University Press.
- Brown, Campbell. 2011. "Consequentialize This." *Ethics* 121: 749–71.
- Christensen, David. 2004. *Putting Logic in Its Place: Formal Constraints on Rational Belief*. Oxford: Oxford University Press.
- Conee, Earl. 2004 [1992]. "The Truth Connection." *Philosophy and Phenomenological Research* 52: 657–69. Reprinted with afterword in Conee and Feldman 2004, 242–58. Page references are to the 2004 reprint.
- Conee, Earl, and Richard Feldman. 2004 [1998]. "The Generality Problem for Reliabilism." *Philosophical Studies* 89: 1–29. Reprinted with afterword in Conee and Feldman 2004, 135–65.
- . 2004 [2001]. "Internalism Defended." In *Epistemology: Internalism and Externalism*, ed. Hilary Kornblith, 231–60. Malden, MA: Blackwell. Reprinted with afterword in Conee and Feldman 2004, 53–82. Page references are to the 2004 reprint.
- . 2004. *Evidentialism: Essays in Epistemology*. Oxford: Oxford University Press.
- . 2008. "Evidence." In *Epistemology: New Essays*, ed. Quentin Smith, 83–104. Oxford: Oxford University Press.
- David, Marian. 2001. "Truth as the Epistemic Goal." In Steup 2001, 151–69.
- . 2005. "Truth as the Primary Epistemic Goal: A Working Hypothesis." In Steup and Sosa 2005, 296–312.
- DePaul, Michael R. 1993. *Balance and Refinement: Beyond Coherence Methods of Moral Inquiry*. London: Routledge.

- . 1998. "Why Bother with Reflective Equilibrium?" In DePaul and Ramsey 1998, 293–309.
- . 2001. "Value Monism in Epistemology." In Steup 2001, 170–83.
- DePaul, Michael R., and William Ramsey, eds. 1998. *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Lanham, MD: Rowman and Littlefield.
- Dewey, John, and James H. Tufts. 1908. *Ethics*. New York: Henry Holt.
- Ellis, Brian. 1988. "Solving the Problem of Induction Using a Values-Based Epistemology." *British Journal for the Philosophy of Science* 39: 141–60.
- Feldman, Richard. 1985. "Reliability and Justification." *Monist* 68: 159–74.
- . 1988. "Epistemic Obligations." *Philosophical Perspectives* 2: 235–56.
- . 2004 [2000]. "The Ethics of Belief." *Philosophy and Phenomenological Research* 60: 667–95. Reprinted in Conee and Feldman 2004, 166–95. Page references are to the 2004 reprint.
- . 2002. "Epistemological Duties." In Moser 2002, 362–84.
- . 2008. "Modest Deontologism in Epistemology." *Synthese* 161: 339–55.
- Feldman, Richard, and Earl Conee. 2002. "Typing Problems." *Philosophy and Phenomenological Research* 65: 98–105.
- Field, Hartry. 1982. "Realism and Relativism." *Journal of Philosophy* 79: 553–67.
- . 1998. "Epistemological Nonfactualism and the A Priority of Logic." *Philosophical Studies* 92: 1–24.
- . 2000. "Apriority as an Evaluative Notion." In *New Essays on the A Priori*, ed. Paul Boghossian and Christopher Peacocke, 117–49. Oxford: Oxford University Press.
- Firth, Roderick. 1998 [1978]. "The Schneck Lectures, Lecture 1: Epistemic Utility." In Firth 1998. First delivered as a lecture at the University of Arizona, 1978.
- . 1998 [1981]. "Epistemic Merit, Intrinsic and Instrumental." *Proceedings and Addresses of the American Philosophical Association* 55: 5–23. Reprinted in Firth 1998, 259–71. Page references are to the 1998 reprint.
- . 1998. *In Defense of Radical Empiricism: Essays and Lectures*, ed. John Troyer. Lanham, MD: Rowman and Littlefield.
- Foley, Richard. 1987. *The Theory of Epistemic Rationality*. Cambridge, MA: Harvard University Press.
- . 1993. *Working without a Net: A Study of Egocentric Epistemology*. Oxford: Oxford University Press.
- Frankena, William K. 1963. *Ethics*. Englewood Cliffs, NJ: Prentice-Hall.
- Fumerton, Richard. 2001. "Epistemic Justification and Normativity." In Steup 2001, 49–60.
- . 2002. "Theories of Justification." In Moser 2002, 204–33.
- . 2006. *Epistemology*. Malden, MA: Blackwell.
- Garvin, Lucius. 1953. *A Modern Introduction to Ethics*. Boston: Houghton Mifflin.

- Gauthier,¹David. 1962. *Practical Reasoning: The Structure and Foundations of Prudential and Moral Arguments and Their Exemplification in Discourse*. Oxford: Oxford University Press.
- Goldman, Alvin I. 1992 [1979]. "What Is Justified Belief?" In *Justification and Knowledge*, ed. George Pappas, 1–23. Dordrecht: Reidel. Reprinted in Goldman 1992, 105–26. Page references are to the 1992 reprint.
- . 1986. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
- . 1992. *Liaisons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge, MA: MIT Press.
- . 1999. *Knowledge in a Social World*. Oxford: Oxford University Press.
- . 2001. "The Unity of the Epistemic Virtues." In *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*, ed. Abrol Fairweather and Linda Zagzebski, 30–48. Oxford: Oxford University Press.
- Goldman, Alvin I., and Erik J. Olsson. 2009. "Reliabilism and the Value of Knowledge." In Haddock, Millar, and Pritchard 2009, 19–41.
- Graham, Peter J. 2011. "Does Justification Aim at Truth?" *Canadian Journal of Philosophy* 41: 51–71.
- Greco, John. 1993. "Virtues and Vices of Virtue Epistemology." *Canadian Journal of Philosophy* 23: 413–32.
- . 1999. "Agent Reliabilism." *Philosophical Perspectives* 13: 273–96.
- Grimm, Stephen R. 2008. "Epistemic Goals and Epistemic Values." *Philosophy and Phenomenological Research* 77: 725–44.
- . 2009. "Epistemic Normativity." In Haddock, Millar, and Pritchard 2009, 243–64.
- Haddock, Adrian, Alan Millar, and Duncan Pritchard, eds. 2009. *Epistemic Value*. Oxford: Oxford University Press.
- Hall, Richard J., and Charles R. Johnson. 1998. "The Epistemic Duty to Seek More Evidence." *American Philosophical Quarterly* 35: 129–39.
- Hazlett, Allan. 2006. "How to Defeat Belief in the External World." *Pacific Philosophy Quarterly* 87: 198–212.
- . n.d. "Limning Structure as an Epistemic Goal." Unpublished manuscript.
- Herman, Barbara. 1993. "Leaving Deontology Behind." In *The Practice of Moral Judgment*, 208–40. Cambridge, MA: Harvard University Press.
- Hieronymi, Pamela. 2006. "Controlling Attitudes." *Pacific Philosophical Quarterly* 87: 45–74.
- Hooker, Brad, Elinor Mason, and Dale E. Miller, eds. 2000. *Morality, Rules, and Consequences: A Critical Reader*. Lanham, MD: Rowman and Littlefield.
- Howard-Snyder, Frances, and Alastair Norcross. 1993. "A Consequentialist Case for Rejecting the Right." *Journal of Philosophical Research* 18: 109–25.
- James, William. 1897. "The Will to Believe." In *The Will to Believe and Other Essays in Popular Philosophy*, 1–31. New York: Dover.

- Jenkins, C. S. 2007. "Entitlement and Rationality." *Synthese* 157: 25–45.
- Kagan, Shelley. 2000. "Evaluative Focal Points." In Hooker, Mason, and Miller 2000, 134–55.
- Kamm, F. M. 1993. *Morality, Mortality*. Vol. 1, *Death and Whom to Save from It*. Oxford: Oxford University Press.
- . 1996. *Morality, Mortality*. Vol. 2, *Rights, Duties, and Status*. Oxford: Oxford University Press.
- Kelly, Thomas. 2003. "Epistemic Rationality as Instrumental Rationality: A Critique." *Philosophy and Phenomenological Research* 66: 612–40.
- . 2008a. "Evidence." In *Stanford Encyclopedia of Philosophy*, fall 2008 ed., ed. Edward N. Zalta. plato.stanford.edu/archives/fall2008/entries/evidence/.
- . 2008b. "Evidence: Fundamental Concepts and the Phenomenal Conception." *Philosophy Compass* 3: 933–55.
- Kornblith, Hilary. 1985. "Ever Since Descartes." *Monist* 68: 264–76.
- . 1993. "Epistemic Normativity." *Synthese* 94: 357–76.
- . 2002. *Knowledge and Its Place in Nature*. Oxford: Oxford University Press.
- Korsgaard, Christine M. 1983. "Two Distinctions in Goodness." *Philosophical Review* 92: 169–95.
- . 1993. "The Reasons We Can Share: An Attack on the Distinction between Agent-Relative and Agent-Neutral Values." *Social Philosophy and Policy* 10: 24–51.
- . 2008. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.
- . 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- Kvanvig, Jonathan. 2003. *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press.
- . 2005. "Truth Is Not the Primary Epistemic Goal." In Steup and Sosa 2005, 285–96.
- . 2011. "Epistemic Justification." In Bernecker and Pritchard 2011, 25–36.
- Lehrer, Keith. 1977. "The Knowledge Cycle." *Nouûs* 11: 17–25.
- . 1981. "Self-Profile." In *Keith Lehrer*, ed. Radu J. Bogdan, 3–104. Dordrecht: Reidel.
- . 1988. "Coherence, Justification, and Chisholm." *Philosophical Perspectives* 2: 125–38.
- Leplin, Jarrett. 2007. "In Defense of Reliabilism." *Philosophical Studies* 134: 31–42.
- . 2009. *A Theory of Epistemic Justification*. Dordrecht: Springer.
- Louden, Robert. 1996. "Toward a Genealogy of 'Deontology.'" *Journal of the History of Philosophy* 34: 571–92.

- Lynch, Michael P. 2004. *True to Life: Why Truth Matters*. Cambridge, MA: MIT Press.
- . 2009a. "Truth, Value, and Epistemic Expressivism." *Philosophy and Phenomenological Research* 79: 76–97.
- . 2009b. "The Values of Truth and the Truth of Values." In Haddock, Millar, and Pritchard 2009, 225–42.
- Maitzen, Stephen. 1995. "Our Errant Epistemic Aim." *Philosophy and Phenomenological Research* 55: 869–76.
- McCloskey, H. J. 1957. "An Examination of Restricted Utilitarianism." *Philosophical Review* 66: 466–85.
- Miscevic, Nenad. 2007. "Virtue-Based Epistemology and the Centrality of Truth (Towards a Strong Virtue-Epistemology)." *Acta Analytica* 22: 239–66.
- Montmarquet, James. 1987. "Epistemic Virtue." *Mind* 96: 482–97.
- . 1993. *Epistemic Virtue and Doxastic Responsibility*. Lanham, MD: Rowman and Littlefield.
- . 2008. "Virtue and Voluntarism." *Synthese* 161: 393–402.
- Moser, Paul K. 1985. *Empirical Justification*. Dordrecht: Reidel.
- , ed. 2002. *The Oxford Handbook of Epistemology*. Oxford: Oxford University Press.
- Muirhead, John H. 1932. *Rule and End in Morals*. London: Oxford University Press.
- Nagel, Thomas. 1970. *The Possibility of Altruism*. Princeton: Princeton University Press.
- Norcross, Alastair. 2006a. "Reasons without Demands: Rethinking Rightness." In *Contemporary Debates in Moral Theory*, ed. James Dreier, 38–53. Malden, MA: Blackwell.
- . 2006b. "The Scalar Approach to Utilitarianism." In *The Blackwell Guide to Mill's Utilitarianism*, ed. Henry R. West, 217–32. Malden, MA: Blackwell.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Percival, Philip. 2002. "Epistemic Consequentialism." *Proceedings of the Aristotelian Society Supplementary Volume* 76: 121–51.
- Pettit, Philip. 1991. "Consequentialism." In *A Companion to Ethics*, ed. Peter Singer, 230–40. Malden, MA: Blackwell.
- . 1997. "The Consequentialist Perspective." In Marcia Baron, Philip Pettit, and Michael Slote, *Three Methods of Ethics: A Debate*, 92–174. Malden, MA: Blackwell.
- Pettit, Philip, and Michael Smith. 2000. "Global Consequentialism." In Hooker, Mason, and Miller 2000, 121–33.
- Piller, Christian. 2009. "Desiring the Truth and Nothing but the Truth." *Noûs* 43: 193–213.
- Pippin, Robert. 2001. "A Mandatory Reading of Kant's Ethics?" *Philosophical Quarterly* 51: 386–93.

- Plantinga, Alvin. 1988. "Positive Epistemic Status and Proper Function." *Philosophical Perspectives* 2: 1–50.
- . 1993a. *Warrant: The Current Debate*. Oxford: Oxford University Press.
- . 1993b. *Warrant and Proper Function*. Oxford: Oxford University Press.
- Portmore, Douglas. 2008. "Dual-Ranking Act-Consequentialism." *Philosophical Studies* 138: 409–27.
- . 2011. "The Teleological Conception of Practical Reasons." *Mind* 120: 117–53.
- Rawls, John. 1963. "Constitutional Liberty and the Concept of Justice." *Nomos* 6: 98–125.
- . 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Reynolds, Steven L. 2002. "Testimony, Knowledge, and Epistemic Goals." *Philosophical Studies* 110: 139–61.
- Ridge, Michael. 2009. "Kantian Consequentialism." *Philosophical Perspectives* 23: 421–38.
- Riggs, Wayne D. 2002. "Beyond Truth and Falsehood: The Real Value of Knowing That *p*." *Philosophical Studies* 107: 87–108.
- . 2003. "Balancing Our Epistemic Goods." *Nous* 37: 342–52.
- . 2008a. "Epistemic Risk and Relativism." *Acta Analytica* 23: 1–8.
- . 2008b. "The Value Turn in Epistemology." In *New Waves in Epistemology*, ed. Vincent F. Hendricks and Duncan Pritchard, 300–323. New York: Palgrave Macmillan.
- Rowbottom, Darrell. 2008. "An Alternative Account of Epistemic Reasons for Action: In Response to Booth." *Grazer Philosophische Studien* 76: 191–98.
- Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- . 2002. "Replies." *Social Theory and Practice* 28: 337–58.
- Scheffler, Samuel, ed. 1988. *Consequentialism and Its Critics*. Oxford: Oxford University Press.
- Schmitt, Frederick. 1992. *Knowledge and Belief*. London: Routledge.
- . 2000. "Veritistic Value." *Social Epistemology* 14: 259–80.
- Shoemaker, Sydney. 1994. "Self-Knowledge and 'Inner Sense', Lecture 2: The Broad Perceptual Model." *Philosophy and Phenomenological Research* 54: 271–90.
- Shogenji, Tomoji. 2012a. "The Degree of Epistemic Justification and the Conjunction Fallacy." *Synthese* 184: 29–48.
- . 2012b. "Internalism and Externalism in Meliorative Epistemology." *Erkenntnis* 76: 59–72.
- Sider, Ted. 1993. "Asymmetry and Self-Sacrifice." *Philosophical Studies* 70: 117–32.
- Slote, Michael. 1985. *Common-Sense Morality and Consequentialism*. Boston: Routledge and Kegan Paul.

- Sosa, Ernest. 1991. *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge: Cambridge University Press.
- Steglich-Petersen, Asbjørn. 2009. "Weighing the Aim of Belief." *Philosophical Studies* 145: 395–405.
- . 2011. "How to Be a Teleologist about Epistemic Reasons." In *Reasons for Belief*, ed. Andrew Reisner and Asbjørn Steglich-Petersen, 13–33. Cambridge: Cambridge University Press.
- . n.d. "Truth as the Aim of Epistemic Justification." Unpublished manuscript.
- Steup, Matthias, ed. 2001. *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*. Oxford: Oxford University Press.
- Steup, Matthias, and Ernest Sosa, eds. 2005. *Contemporary Debates in Epistemology*. Malden, MA: Blackwell.
- Thomson, Judith Jarvis. 1976. "Killing, Letting Die, and the Trolley Problem." *Monist* 59: 204–17.
- Tsanoff, Radoslav A. 1947. *Ethics*. New York: Harper and Brothers.
- Vahid, Hamid. 2006. "Aiming at Truth: Doxastic vs. Epistemic Goals." *Philosophical Studies* 131: 303–35.
- . 2010. "Rationalizing Beliefs: Evidential vs. Pragmatic Reasons." *Synthese* 176: 447–62.
- . 2011. "The Concept of Entitlement and Its Epistemic Relevance." *Pacific Philosophy Quarterly* 92: 380–99.
- Velleman, David. 2000a. "On the Aim of Belief." In Velleman 2000b, 244–81.
- . 2000b. *The Possibility of Practical Reason*. Oxford: Oxford University Press.
- . 2009. *How We Get Along*. Cambridge: Cambridge University Press.
- Wedgwood, Ralph. 2002. "The Aim of Belief." *Philosophical Perspectives* 16: 267–97.
- . 2008. "Contextualism about Justified Belief." *Philosophers' Imprint* 8, no. 9: 1–20.
- Weinberg, Jonathan M. 2007. "Moderate Epistemic Relativism and Our Epistemic Goals." *Episteme* 4: 66–92.
- Whitcomb, Dennis. 2007. "An Epistemic Value Theory." PhD diss., Rutgers University.
- . 2011. "Wisdom." In Bernecker and Pritchard 2011, 95–105.
- Willenken, Tim. 2012. "Deontic Cycling and the Structure of Commonsense Morality." *Ethics* 122: 545–61.
- Williams, Bernard. 1973a [1970]. "Deciding to Believe." In *Language, Belief, and Metaphysics*, ed. Howard E. Kiefer, and Milton K. Munitz, 95–111. Albany: State University of NY Press. Reprinted in Williams 1973a, 136–51. Page references are to the 1973 reprint.
- . 1973a. *Problems of the Self: Philosophical Papers 1956–1972*. Cambridge: Cambridge University Press.

- . 1973b. "A Critique of Utilitarianism." In J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against*, 75–150. Cambridge: Cambridge University Press.
- Zimmerman, Michael J. 2002. "Taking Luck Seriously." *Journal of Philosophy* 99: 553–76.
- Zink, Sidney. 1962. *The Concepts of Ethics*. London: MacMillan and Co.

Peer Disagreement, Evidence, and Well-Groundedness

Han van Wietmarschen

Stanford University

1. Introduction

The peer disagreement debate is concerned with cases in which you have good reason to believe your epistemic peer disagrees with you. The central question in this debate is formulated in a number of different ways. What should you believe about the disputed proposition? What would it be reasonable for you to believe? What would you be justified to believe? How confident should you be? What degree of confidence would be rational for you? In this article, I show that these questions are ambiguous between two different standards of epistemic justification. We can understand these questions to be exclusively about evidential support. On this understanding, we ask which doxastic attitude is supported by your evidence once you learn about the disagreement. Alternatively, we can understand the question to be about well-grounded belief. On this reading, we don't just consider the evidence provided by the disagreement, we also look at possible effects of the disagreement on the well-groundedness of your beliefs.

Special thanks to Krista Lawlor for ideas, criticism, and encouragement throughout this project. I would also like to thank Michael Bratman, Anna-Sara Malmgren, and an anonymous referee for the *Philosophical Review* for helpful discussions of these ideas or comments on earlier drafts. I benefited from discussing this material, in one form or another, with fellow students in Stanford's dissertation development seminar, the students in my seminar on reasonable disagreement at Leiden University, and audience members at the Leiden University practical philosophy colloquium. Thanks to Lauren Quan for preventing many minor mistakes.

I argue that keeping this distinction in view has important implications for the peer disagreement debate. I focus on so-called “conciliatory views.” According to these views, in peer disagreement cases generally, you should suspend judgment on the disputed issue, or be significantly less confident in your view. Understandably, conciliationism has been the focus of much of the peer disagreement debate: it is a powerful view with potentially sweeping consequences for the justification of our controversial beliefs. I first argue that for a wide range of conceptions of evidential support, conciliatory views fail if we understand them exclusively in terms of evidential support (section 3). This leaves conciliatory views with two strategies: appeal to an alternative standard of evidential support, or appeal to nonevidential considerations. I argue that the first strategy faces serious difficulties (section 4). I then go on to defend a conciliatory view of peer disagreement as a view about well-grounded belief (section 5). In peer disagreement cases generally, you are required to suspend judgment on the disputed proposition because once you learn of the disagreement your belief is no longer well grounded.

2. Conciliatory Views of Peer Disagreement

The debate on peer disagreement is concerned with what people should believe when they find themselves in a particular kind of disagreement. Here is a much discussed example, call it “Bill Calculation”:

You and S went out for dinner, and it is time to pay the bill. You and S always split the bill equally and tip 20 percent. You have gone out for dinner with S often over the course of many years. You always both do the calculation in your head and compare answers. In the past, whenever you disagreed about your shares, you and S each have been wrong roughly half the time. You and S are presented with the bill, you can both clearly see that the total is \$46. You and S both do the calculation in your heads, you believe that your shares are \$27.60. S says your shares are \$27.10 (Comparable to Christensen 2007, 193).

As it happens, you are correct: your shares are \$27.60. The basic features of this disagreement, and peer disagreements generally, are as follows. You and some other person S disagree about some proposition *p*. This can be cashed out in terms of flat-out attitudes or in terms of degrees of credence. I will speak in terms of flat-out attitudes (belief, disbelief, and suspension of judgment) because that simplifies the discussion in a number of ways. I will assume that in cases of peer disagreement, you

believe p , and S believes $\sim p$.¹ In the debated cases, you not only know that S disagrees with you; you are also justified to believe that S is your *epistemic peer*. The notion of “epistemic peerage” has two components. First, you are justified to believe that S has the same evidence about p as you do. Second, you are justified to believe that S is equally good at reasoning about p as you are.² There are many questions about how the notion of peerage should be understood exactly, but I will leave these issues aside.³

It will be useful to think of peer disagreement cases as conforming to the following schema; in parentheses are the corresponding features in the bill calculation case:

- t_0 : You have no attitude toward p , but you are justified to believe that S is equally good at reasoning about p (you have evidence that you have an equally good track record when it comes to bill calculation).
- t_1 : You gain access to some evidence about p , and you now have evidence E about p . You are justified to believe that S also has evidence E about p (you and S learn that the total is forty-six dollars, you also both know that you split the bill equally and that you tip 20 percent).
- t_2 : You form the belief that p after a process of reasoning on the basis of E . You are justified to believe that S also formed a belief about p on the basis of a process of reasoning on the basis of E (after calculating your shares in your head, you conclude that your shares are \$27.60, S seems to be doing a similar calculation).

1. Kelly (2010, 117–18) argues that we ultimately want an account of the epistemology of peer disagreement in terms of degrees of confidence to be able to account for disagreements among more than two people and disagreements between believers and agnostics. I agree with Kelly, but these cases will not matter for my discussion.

2. We can think of this equal competence in terms of equality in general epistemic virtues relevant to issues like p , such as thoroughness, intelligence, and perspicacity. Gutting (1982, 83) uses the concept in this way and others follow him (see Kelly 2005 and Christensen 2007). Alternatively, we can think of this second aspect in terms of equal reliability (see Enoch 2010 and Elga 2007). This second alternative gives rise to some interesting questions about how reliability should be understood (see Lam 2011).

3. Some authors understand peerage entirely in terms of reliability: you and S 's being epistemic peers about issues like p just means that you and S are equally likely to be right about issues like p . On that view, you and S could be in a peer disagreement even when the formation of your beliefs does not involve a process of reasoning at all. For example, Elga (2007, 486) describes a case in which you and S both see the finish of a horse race and disagree about which horse won. As I understand cases of peer disagreement, they involve conflicting views arrived at after a process of reasoning given the same evidence.

t_3 : You learn that S believes $\sim p$ (S tells you that she thinks your shares are \$27.10).

The key question of the peer disagreement debate is: which doxastic attitude toward p is justified for you at t_3 ? Conciliatory views hold that, in peer disagreement cases generally, you should suspend judgment on p (see Feldman 2007a, 2007b). Or, more commonly, they hold that you should, at t_3 , revise your degree of confidence significantly in the direction of S's view (see Christensen 2007, 2010b; and Kornblith 2010).⁴ As I mentioned, my discussion will be stated in terms of flat-out attitudes. This is unproblematic because my discussion will be about the basic motivating ideas behind conciliatory views, and these are the same regardless of whether the view is stated in terms of flat-out attitudes or degrees of credence.

Why do conciliationists think that you should suspend judgment on p ? I will rely on David Christensen's work to state the conciliatory position because I think his is the most developed and best-defended view of this kind. I will make changes where necessary to accommodate talk of flat-out attitudes. Christensen's argument rests on two key principles that apply to cases of disagreement generally but that lead to a conciliatory view in cases of peer disagreement.

The first principle relates a subject's justification for the belief that p to his or her justification to believe various explanations of his or her disagreement about p . Conciliationists hold that evidence that S disagrees about p provides you with evidence that someone made a mistake in reasoning about p . For this to be a relevant observation, it had better be the case that what you are justified to believe about whose mistake explains the disagreement is relevant to what you are justified to believe about p . As Christensen (2010a) shows, it is surprisingly hard to spell out general principles that relate the justification of lower-order beliefs to the justification of higher-order beliefs. At least for now, the following will do:

Integration: a subject is not justified to believe that p if: (a) he or she believes p , and he or she is justified to believe that some other subject, S, believes $\sim p$, (b) he or she is justified to believe that either his or her own or S's mistake explains his or her disagreement with S about p , and (c) he or she is not

4. One important example of a conciliatory view is the "equal weight view," according to which you should, at t_3 , have a degree of confidence in the disputed proposition that is the average of the degrees of confidence you and S had at t_2 (see Elga 2007).

justified to believe that S's mistake, rather than his or her own mistake, explains his or her disagreement about p.

Note that the principle does not entail the claim that you are justified to believe p only if you are justified to believe you did not make a mistake in reasoning about p.

The second key idea is that in evaluating the justification of beliefs about the various explanations of your disagreement with S, we should bracket or disregard your original reasoning on the disputed proposition. Christensen takes this idea to be at the heart of his conciliatory view, and he calls it the "independence principle." As we will see below, the details of this principle will be important. For now, I will state it as follows:

Independence: when we determine what a subject is justified to believe about the explanation of his or her disagreement with S about p, we should bracket the subject's reasoning about p.

This principle rules out a response to the peer disagreement problem that seems problematically circular. Suppose we accept Integration so your justification to believe that p depends on whether you are justified to believe that S, rather than you, made a mistake. We then say: you engaged in a process of reasoning about p and concluded that p. Since S concluded $\sim p$, this justifies you to believe that S made a mistake. Since you are justified to believe that S made a mistake, S's disagreement does not threaten the justification for your belief that p. This seems problematic because your reasoning about p is challenged by S's disagreement, and we appeal to that very process of reasoning to establish that S must have made a mistake.

With these principles in hand, the argument for conciliationism is fairly easy to state. In schematic form it goes like this:

- (1) We assume (a): you believe p, and you are justified to believe S believes $\sim p$.
- (2) We assume (b): you are justified to believe that your disagreement with S is explained by a mistake on either your own or S's side.
- (3) By hypothesis, you are justified to believe that S is your epistemic peer, so S is equally informed about the issue under dispute and is equally good at performing the kind of reasoning involved.

- (4) From (3) and Independence, we get (c): you are not justified to believe that S's mistake rather than your own explains your disagreement with S.
- (5) from (a), (b), (c), and Integration, we get: you are not justified to believe that p.⁵

Both Integration and Independence have considerable intuitive appeal, so I think that if this argument can be made to work, it provides a powerful case for conciliationism. Steps (1) and (3) are straightforward, and I will not question (2).⁶ I will also assume that Integration is correct, though we will see below that the version of conciliationism I want to defend needs a different statement of this principle.⁷ This leaves (4) as the key step: the independence principle together with your justification for believing that S is your epistemic peer gives us the conclusion that, at t_3 , you are not justified to believe that S's mistake, rather than your own, explains the disagreement.

5. On this reconstruction of the argument, conciliationism holds that at t_3 you are not justified to believe that p. Depending on one's conception of suspension of judgment, this may fall short of the conclusion that you should, at t_3 , suspend judgment on p. I choose this negative formulation to avoid debates about the proper interpretation of suspension of judgment and because I think the key claim all versions of conciliationism share is that peer disagreement undermines the justification of our views, not that peer disagreement justifies some particular attitude. I will return to the possible limitations of this approach in section 5.

6. The notion of a "mistake" does not refer to a false belief but to a rationally sub-optimal belief. This means that assuming (b) is important for Christensen because it allows him to treat the issue of peer disagreement as a special case of a broader class of situations in which you acquire evidence of your own rational failure (see Christensen 2010b). I think (b) does not hold in an interesting class of (peer) disagreement cases, including important cases of philosophical, religious, moral, and political disagreement. I discuss these issues in van Wietmarschen n.d.

7. You may wonder why we need Integration and the detour through higher-order beliefs at all: why not apply the independence principle directly to the belief that p? Let Independence require that when we evaluate what you are justified to believe about p, in cases in which you are justified to believe that S disagrees with you about p, we bracket your reasoning about p. Having bracketed your reasoning, you are not justified to believe p, which is what the conciliatory view needs. The problem is that Independence would yield this answer in all cases of disagreement, even when you have reason to think S is your epistemic inferior. Integration is needed to let the evidence that S is your epistemic peer play an appropriate role.

3. Evidential Support and Conciliationism

Can we take step (4)? Are you, at t_3 , not justified to believe that S's mistake, rather than your own, explains your disagreement with S? The answer will depend on how we understand the notion of being justified to believe something. Peer disagreement cases are usually discussed exclusively in terms of the evidence available to the subject. This suggests that "being justified to believe" is also understood to be exclusively a matter of the available evidence. In this section, I will assume that S is justified to believe that p , when E is the evidence available to S, if and only if E supports the belief that p . So our question becomes: does the independence principle together with the evidence that S is your epistemic peer lead to the conclusion that, at t_3 , your evidence does not support the belief that S's mistake explains the disagreement? In this section, I argue that a broad range of interpretations of the notion of evidential support lead to a negative answer. As a consequence, conciliatory views are false if we understand them in terms of a notion of evidential support that falls in this range.

To introduce this range of conceptions of evidential support, here is an example:

Evidential Support (ES): A body of evidence E supports the belief that p for subject S if and only if S would believe that p on ideal reflection on whether p , given access to E .

This conception is inspired by Richard Foley's conception of epistemic rationality (see Foley 1987, 1993). There is much to say about this principle, but one key feature will be central to my discussion. To illustrate this feature, look at t_1 and t_2 in the bill calculation case, before there is any disagreement. At t_1 your evidence about your shares of the bill is roughly this: the total is \$46, you share the bill equally, and you tip 20 percent. What does this evidence support according to ES? On ideal reflection, given access to this evidence, you would believe that your shares are \$27.60. So, according to ES, your evidence at t_1 supports the belief that your shares are \$27.60. Notice that this claim is true even before you did any reasoning about your shares of the bill and before you reached a conclusion on the basis of that reasoning. For t_2 , after you engaged in a process of reasoning and reached a conclusion, ES simply

says the same thing: your evidence supports the belief that your shares are \$27.60.⁸

These observations about the bill calculation case illustrate that ES makes the following claim false:

Required Reasoning (RR): in cases where subject S has access to evidence E but it takes a process of reasoning to *see* that E provides evidence for p, E only in fact supports the belief that p for subject S if S engaged in the right kind of reasoning process.

The thought is that when you learn that the total is \$46, and you know you tip 20 percent and share equally, you have evidence that indicates that your shares are \$27.60, but unless you are some kind of savant you cannot *tell* right away that your evidence supports this conclusion. You need to think through the calculation to get there. ES implies that your actually going through this reasoning is not necessary for your evidence to support the belief that your shares are \$27.60.

I think there is much to be said for a Foley-style interpretation of evidential support. But many readers may, for various reasons, reject this interpretation. Fortunately, my arguments will apply to any interpretation of evidential support that makes RR false, and it seems that there is a broad range of conceptions of evidential support that do so. Think of views on which evidence E supports the belief that p just when E makes p sufficiently probable, or just when E confirms p, or just when the belief that p fits E.⁹ I will state my arguments in terms of ES, but the specific Foley-inspired features of this principle will not be essential to the argument.¹⁰

8. ES says that the evidence supports this belief even if you do not in fact have the belief. In this sense, ES is a standard of what you are justified *to* believe, not what you are justified *in* believing. This is sometimes referred to as the distinction between propositional and doxastic justification. I will say more about this distinction below.

9. At least given natural interpretations of probabilification, confirmation, and “fitting.” There are no doubt also ways to spell out these ideas that make RR true.

10. One feature of ES is especially relevant for the peer disagreement debate. Depending on how we understand “ideal reflection,” whether some body of evidence supports the belief that p may be relative to the subject. This is a controversial feature and a violation of what is known in the peer disagreement debate as the “uniqueness thesis,” the thesis that any body of evidence justifies a unique doxastic attitude, regardless of the subject who has access to that body of evidence. The uniqueness thesis plays an interesting role in the peer disagreement debate: there is both disagreement about whether the thesis

In the rest of this section, I argue that conciliatory views fail if we interpret them in terms of ES, or any other conception of evidential support that makes RR false. The conciliatory argument rests on the claim that, at t_3 , your evidence does not support the belief that S's mistake explains the disagreement. At t_3 , the total evidence available to you is roughly this:

- The total bill is \$46, you share equally and tip 20 percent
- After a quick mental calculation, S thought that your shares are \$27.10, and you thought your shares are \$27.60
- All the evidence that supports the claim that S is your epistemic peer

Call this set of evidence E_{t_3} . If we ignore the independence principle, ES leads to the conclusion that E_{t_3} *does* support the belief that S's mistake explains the disagreement. Imagine that you engage in an ideal process of reflection given E_{t_3} . When we imagine an ideal process of reflection we abstract away from a range of features of regular, nonideal reflection. We imagine that there are no limitations on our concentration and on the time spent thinking about the issue, and we imagine that there are no distorting factors such as intoxication, fatigue, or carelessness.¹¹ Ideal reflection on the original evidence would lead you to conclude that your shares are \$27.60. How would you, on ideal reflection, weigh the evidence provided by the disagreement of your epistemic peer? It seems that from the vantage point of ideal reflection on all the evidence, the fact that S thought, after a quick mental calculation, that your shares are \$27.10 would not lead you to change your mind. Rather, you would conclude that S must have made a calculation mistake and that this mistake explains your disagreement.

The key idea is that in cases like Bill Calculation, you and S disagree after processes of reasoning that are in various ways imperfect: you and S both did a simple calculation in your heads, you did not take

is true (see White 2005 and Feldman 2007b) and about whether conciliatory views depend on its truth (see Christensen 2007, 2009; Kelly 2010; Goldman 2010; Ballantyne and Coffman 2012). None of my arguments below depend on the fact that ES violates the uniqueness thesis.

11. These idealizations are supported by intuitions about evidential support: it doesn't seem that how much time I have to think about some issue, or how distracted or intoxicated I am, affects what evidence is available to me and what that evidence supports. Some authors dispute this picture. I say more about such views in section 4.

a lot of time to do it, you weren't particularly diligent, and the quality of your reasoning processes may have suffered from certain distractions, tiredness, and so on. So from the vantage point of ideal reflection on all the evidence, including the original evidence, the fact that S disagrees after such a nonideal process of reflection does not carry very much weight.¹² Of course, from the perspective of ideal reflection, your own actual, nonideal process of reflection does not carry much weight either, but that is exactly what we would expect from a standard of evidential support that makes RR false.

If E_{t_3} does not support the belief that S's mistake explains the disagreement, it must be because of the independence principle. To see whether the independence principle leads to a different conclusion, we need to look at this principle in more detail. Here is one formulation from Christensen (2011, 1):

Independence (1): In evaluating the epistemic credentials of another's expressed belief about p, in order to determine how (or whether) to modify my own belief about p, I should do so in a way that doesn't rely on the reasoning behind my initial belief about p.

12. Could we resist the argument by insisting on a different interpretation of the idealization involved in ES? This question points to a large issue that I can give only cursory treatment here. My argument exploits the gap between ideal reflection and the kind of reflection we perform in cases like Bill Calculation. To avoid the argument, we would need an interpretation of idealization on which this gap closes. One sense in which your actual process of reflection was ideal is that it led you to exactly the answer that was in fact supported by your evidence. But this cannot be the understanding of idealization at work in ES because it would make ES vacuous: it would say that what your evidence supports, at t_3 , is just the belief that you would form on reflection that leads to the belief that is in fact supported by your evidence. Your reflection was also ideal, we might say, in that you did not make any mistakes. But if ideal reflection means reflection that is free from mistakes, this would make ES into an indeterminate standard: there are many different mistake-free processes of reasoning that would lead to a conclusion about your shares of the bill, some quick and careless and some protracted and diligent, and these different processes lead to different conclusions about evidential support. Which do we choose? We cannot take your actual process of reasoning as the starting point and take idealization to involve correcting for mistakes in that process, because then it is unclear how ES would make RR false: before you did any reasoning, there is no actual process of reflection to correct for mistakes. These remarks indicate that two candidate conceptions of ideal reflection that count your actual process of reflection in Bill Calculation as ideal face some serious difficulties. This does not definitively establish that there could not be a suitable conception of ideal reflection that would avoid my argument, but it does indicate that such a conception is not easy to come by. Thanks to an anonymous reviewer from the *Philosophical Review* for pressing me on this issue.

Ignore some of the differences between this and my earlier statement of the principle. The key point is that according to Independence (1), it is clear that *what* we should bracket, or not rely on, is the actual process of reasoning you engaged in between t_1 and t_2 : the principle asks you not to rely on the reasoning that led to your belief in the first place. Call this the “restricted interpretation” of the independence principle.

On the restricted interpretation, the principle does not lead to a conciliatory view. We ask: would you, on ideal reflection, given all the evidence you have at t_3 , believe that S’s having made a mistake explains the disagreement? When we answer this question, we make sure to bracket your original, actual process of reasoning. This does not make a significant difference. Careful reflection on the original evidence would still lead you to conclude that your shares are \$27.60. Since S’s conflicting belief rests on a quick mental calculation, you would, on ideal reflection, conclude that S must have made a mistake. You do not at any point need to appeal to your original reasoning to reach those conclusions. Given ES, the evidence available to you at t_3 supports the belief that S’s mistake explains the disagreement, even if we accept the independence principle. So at t_3 you are justified to believe S’s mistake, rather than your own, explains the disagreement. This blocks step (4) in the conciliatory argument.

A natural worry about this line of thinking is that when assessing what you would, on ideal reflection, believe to be the explanation of your disagreement with S, I imagine you to reflect on the issue under dispute. I imagine the ideal reasoning to involve calculating your shares based on the original evidence, and I imagine you to then assess what explains your disagreement with S after you both did a quick mental calculation. Perhaps this is in conflict with a proper understanding of the independence principle. Another of Christensen’s (2007, 199) formulation of the independence principle could support this idea:

Independence (2): I should assess explanations for the disagreement in a way that’s independent of the reasoning on the matter under dispute.

Independence (2) is ambiguous between a requirement to bracket your original process of reasoning and a requirement to bracket any and all reasoning on the matter under dispute. Call the latter the “unrestricted interpretation.”

The unrestricted interpretation does lead to a conciliatory view. Would you, on ideal reflection, believe that your disagreement with S is

explained by S's having made a mistake? You take into consideration all the evidence you have available to you at t_3 , but we make sure to bracket any reasoning at all about what your shares of the bill are. This means that evidence that the total is \$46, that you tip 20 percent, and that you share the bill equally is irrelevant: if you do not reflect on this evidence at all, then this evidence does not lead you to believe anything in particular about whose mistake explains the disagreement. So it seems that the only remaining evidence is all the evidence that suggests that S is your epistemic peer and the fact that you believe that your shares are \$27.60 and that S believes they are \$27.10. On ideal reflection given just this evidence, it seems that you would consider you and S to be about equally likely to be mistaken. So you would not believe that S's mistake, rather than your own, explains the disagreement. This gives us what we need for step (4): given ES, you are not justified, at t_3 , to believe that S's mistake, rather than your own, explains the disagreement.

How to choose between the restricted and the unrestricted reading of the independence principle? The motivation for the independence principle is that it seems circular to dismiss the view of your disagreeing peer as mistaken on the basis of the very reasoning that is under dispute (see Christensen 2007, 198–99; 2011, 2). But now we can ask: what exactly does the disagreement with your peer, in the bill calculation case, call into question? Does it tell you that you are not capable of doing the kind of calculations required to figure out your shares of the bill? Does it provide you with evidence that you are incompetent when it comes to division and addition, or that you don't understand percentages? Without claiming that your disagreeing peer does not provide you with *any* evidence for these claims, I think it is safe to say that the answer is “no.” On the most natural reading of the case, the fact that your peer disagrees with you does not call into question your ability to reason through relatively simple calculation problems, nor does it call into question S's ability to do so.

Rather, the disagreement with your peer indicates that something went wrong in either your own or S's actual processes of reasoning. The most plausible explanation of the disagreement is that *in this particular case*, either you or S made a mistake. Intuitively, it would be problematic to dismiss the view of your peer as mistaken if you rely on the actual reasoning process you engaged in between t_1 and t_2 to do so but not if you rely on your general competence at simple mathematical reasoning. The disagreement in the simple calculation case calls into question neither your own nor S's status as a generally reliable calculator of your shares of

the bill. All this indicates that the independence principle should be given the restrictive rather than the unrestrictive interpretation.¹³

I just argued that the independence principle should be interpreted in a restrictive sense. This interpretation, together with ES, leads to the conclusion that at t_3 , your evidence supports the belief that S's mistake explains the disagreement. So, if we understand epistemic justification in terms of evidential support, and evidential support in terms of ES, the conciliatory argument fails at step (4).

My argument shows that conciliatory views are false as views about evidential support, because in Bill Calculation your evidence supports the belief that your shares are \$27.60, despite your disagreement with S. The argument does not show, however, that the disagreement of our epistemic peers never provides us with strong evidence against our views. S's disagreement does not provide you with very powerful evidence because the actual processes of reasoning you and S engaged in were relatively quick and careless. But when the processes of reasoning that lead to the disagreement are much closer to ideal processes of reflection, you would, even on ideal reflection, take S's disagreement very seriously as evidence about the disputed issue.¹⁴

13. Christensen's (2010b, 193–99) discussion suggests a third possible interpretation of the independence principle: independence requires that we bracket your original *evidence*, rather than your reasoning, actual or otherwise. This would also lead to a conciliatory view about evidential support. This interpretation suffers from problems similar to those facing the unrestricted interpretation. In the simple calculation case, and many other cases of peer disagreement, it is not true that the disagreement indicates that there is something wrong with your original evidence. Again, the disagreement may provide *some* evidence for this claim, but the disagreement does not call into question that the total is forty-six dollars, that you tip 20 percent, and that you share equally. Moreover, this reading runs into an additional problem in cases of disagreement where a priori reasoning would make it very likely that your original answer is correct. Take a case like Bill Calculation, but let the disputed proposition be "eighty-nine is a prime number." At t_3 , you believe eighty-nine is a prime number; S disagrees. Even if we bracket all your evidence, you would still conclude, on ideal reflection, that eighty-nine is a prime number and that S must have made a mistake. It would be strange if conciliationism would hold that peer disagreement requires you to significantly revise your view, except in such a priori cases.

14. In van Wietmarschen n.d., I argue that some of our profound and persistent peer disagreements about religion, philosophy, and politics have a special epistemic significance that is not captured by general conciliatory views. In contrast to cases like Bill Calculation, such disagreements provide us with powerful evidence against our view. So while I think all peer disagreements are epistemically similar in that they all undermine the well-groundedness of our beliefs (see section 5), I believe that they are epistemically dissimilar with regard to the evidence they provide us with. As far as the evidence is concerned, my view shows similarities to the views of Feldman (2009) and Kelly (2010).

My argument generalizes to any conception of evidential support that makes RR false. As we saw, the independence principle is essential to the conciliatory argument. I argued that the independence principle should be interpreted in a restrictive sense, requiring us to bracket your actual process of reasoning. Together with ES, this does not lead to a conciliatory view because this actual process of reasoning isn't necessary for your belief that your shares are \$27.60 to be supported by your evidence in the first place. Consequently, bracketing this actual process of reasoning doesn't have major consequences for what you are justified to believe about your shares once you find out about S's disagreement. But according to any conception of evidential support that makes RR false, your actual reasoning is not necessary for evidential support. So for all those views, the independence principle requires us to bracket something that isn't needed for your belief that your shares are \$27.60 to be supported by your evidence. As a result, the independence principle does not lead to a conciliatory view about any such conception of evidential support.

The peculiar features of ES are not necessary to reach this conclusion. Even if we do not think of evidential support in terms of idealized reflection, we do not reach conciliatory answers about evidential support. Think of views on which evidence E supports proposition p insofar as E confirms p, or insofar as E makes p probable. On such views, it is even more obvious that conciliatory views are false: it is hard to deny that the evidence available to you at t_3 confirms or makes probable the conclusion that your shares are \$27.60. After all, that conclusion is entailed by your evidence, and bracketing your actual reasoning process will not change that fact.¹⁵

15. Does this beg the question against conciliationism? After all, conciliatory arguments point to the relevance of evidence about your own rational failure, and it seems that ES and other conceptions of evidential support that make RR false simply abstract away from such evidence. The answer is that these conceptions of evidential support do not abstract away from the evidence S's disagreement provides about the imperfections of your own actual reasoning. The point is, rather, that these conceptions of evidential support do not consider the evidence provided by S's disagreement to be very weighty compared to the original evidence. In this regard, my view is similar to Kelly's (2010). To see that the evidence S's disagreement provides about our own rational failure is not simply dismissed, consider what would happen if you had lost your original evidence. In that case, all you have to go on is your own belief that your shares are \$27.60, S's belief that they are \$27.10, and all the evidence that indicates you and S are peers. ES would lead to the conclusion that you are not justified to have either view: ideal reflection on just that evidence would lead you to suspend judgment on what your shares are. The same

I conclude that conciliatory views fail if we interpret them in terms of ES, or in terms of any conception of evidential support that makes RR false. What is the significance of this conclusion? First, we have an interest in how well people's beliefs are supported by their evidence: we criticize other people's beliefs when they are not supported by their evidence, and we worry whether we have sufficient evidence for what we believe ourselves. So it is important in its own right to know that conciliatory views are not true insofar as we are concerned with such an epistemic standard.

Second, the discussion in this section provides us with a way to interpret some of the views in the literature that compete with conciliationism. In particular, Kelly (2010, 135–50) argues for a view he calls the “total evidence view.” On this view, what you are justified to believe in peer disagreement cases is determined entirely by the total evidence available to you, including the original evidence (if you retained it) and evidence provided by the disagreement. Though in some cases the total evidence supports suspension of judgment, Kelly thinks that conciliationism is false because in many cases the total evidence supports (roughly) your original belief or the belief of your disagreeing peer. These claims are true if we understand the peer disagreement problem in terms of ES or a standard like it. This at the very least suggests that we might interpret Kelly's view as one about a standard of evidential support that makes RR false.

4. Reasoning and Evidence

In section 3 we saw that conciliationism fails if we interpret it in terms of standards of evidential support that make RR false. This cannot be the end of the story about conciliationism. First, there is undeniably a set of conciliatory intuitions about peer disagreement that we need to account for in some way: it seems that in the bill calculation case, there is some sense in which you are not justified, at t_3 , to believe that your shares are \$27.60. Second, it is fairly obvious that your reasoning between t_1 and t_2 is epistemically significant, though not for ES and certain other standards of evidential support. According to ES, you are, at t_1 , justified to believe that your shares are \$27.60. But if you were to believe, at t_1 , before you did any reasoning, that your shares are \$27.60, then there would be

conclusion could be supported by other conceptions of evidential support, even if they make RR false.

something very problematic about this belief. The belief seems to be a mere lucky guess.

At this point, it may help orient the reader to mention some terminology. Intuitively, there is a sense in which you are, at t_1 , justified to believe that your shares are \$27.60, but there is also a sense in which you are not. People have talked about this distinction in various ways. Some would hold that the belief is *justifiable* for you but not *justified* (see Korcuska 1997, 171), or that you are justified *to* believe that your shares are \$27.60 but that you are not justified *in* believing it. This is sometimes talked about in terms of two different objects of evaluation: the proposition that your shares are \$27.60 is epistemically justified for you, but the belief that they are is not. Others also draw a distinction between propositional and doxastic justification but apply both to beliefs: the belief that your shares are \$27.60 is propositionally but not doxastically justified (see Foley 1987, 175–86). For my purposes, these terminological differences do not matter. What matters is that the widespread use of this distinction suggests that even if there are legitimate epistemic standards, such as ES, that are insensitive to facts about your reasoning, there are also important standards of epistemic justification that are sensitive to such facts.

In this section and the next, I discuss two ways to draw on these intuitions in support of conciliationism. The first is to take the intuition that you are not, at t_1 , justified to believe that your shares are \$27.60 to indicate that ES, and other principles that make RR false, represents too narrow a conception of evidence. On a more expansive notion of evidence, the belief that your shares are \$27.60 is not supported by your evidence until you go through the required reasoning. I discuss this proposal in this section and will argue that it faces some serious problems. The second is to take the intuition to track considerations of well-groundedness or proper basing. On such a view, the belief that your shares are \$27.60 is supported by your evidence at t_1 , but without going through the reasoning, your belief is not well grounded in your evidence. In section 5, I argue in favor of a conciliatory view in terms of well-groundedness.

In the rest of this section, I discuss the proposal to think of conciliationism in terms of a standard of evidential support that makes RR true. According to such a standard, the reasoning process in which you engage between t_1 and t_2 provides you with evidence about p , and this evidence is necessary for you to be justified to believe p . I will call the view that appeals to such a standard “Reasoning as Evidence.” Reasoning as Evidence supports step (4) in the argument for conciliationism. In Bill

Calculation, at t_1 your evidence does not support the belief that your shares are \$27.60; at t_2 you acquire new evidence due to your reasoning such that your evidence does support the belief that your shares are \$27.60. We ask: are you, at t_3 , justified to believe that S's mistake explains the disagreement? The independence principle asks us to bracket your actual reasoning process about the issue under dispute. This means that we have to bracket the evidence generated by your reasoning, which makes your original evidence basically irrelevant: as was the case at t_1 , it does not support the belief that your shares are \$27.60, so it does not tell you anything about who made the mistake. The rest of your evidence (your own and S's beliefs and the evidence that S is your epistemic peer) does not, by itself, support the belief that S's mistake, rather than your own, explains the disagreement. This gives us what we need for step (4) in the conciliatory argument.

Your reasoning between t_1 and t_2 can be taken to generate evidence about p in two ways. First, one might think that, at t_2 , you end up with higher-order evidence for p , constituted by evidence that you just completed a process of reasoning by concluding that p , and background evidence of your own reliability as a reasoner. Second, one might think that your reasoning provides evidence by adding the conclusions of the inferential steps you take to your evidence. In Bill Calculation, you start out with evidence that the total is forty-six dollars, you tip 20 percent, and you share equally. Suppose you start by thinking that one-tenth of forty-six is 4.6. We add "one-tenth of forty-six is 4.6" to your evidence. You then think that 4.6 times two is 9.2, so we add "4.6 times two is 9.2" to your evidence, and so on. We could call this "linking evidence," after Matheson 2009. Suggestions in the direction of Reasoning as Evidence can be found in Feldman and Conee 2008 and explicitly as a view about peer disagreement in Matheson 2009. If this picture succeeds, then conciliationism can be defended in purely evidentialist terms, as long as we have the right conception of evidential support in mind.¹⁶ This conclusion would be especially welcome to those with evidentialist leanings.

Reasoning as Evidence faces some serious problems. First, Reasoning as Evidence leads to counterintuitive claims about what our original

16. Note that Reasoning as Evidence needs not just the claim that your reasoning about p can provide you with *some* evidence relevant to p . Rather, it needs the claim that this evidence is *necessary* for you to have a justified belief that p , even if you retained your original evidence. This is a strong claim to be committed to, and I think there is a considerable burden on defenders of this view to argue for it.

evidence supports. In any case in which justified belief about p requires a process of reasoning, Reasoning as Evidence says that the original evidence, on the basis of which you do your reasoning, does not support any doxastic attitude toward p , or supports suspending judgment about p . In Bill Calculation, your original evidence does not, strictly speaking, support the \$27.60 conclusion. This is a counterintuitive result: it seems that the original evidence supports the belief that your shares are \$27.60. It doesn't seem correct to say that the evidence available to you does not support any conclusion about your shares of the bill, or that it supports suspension of judgment.

Second, Reasoning as Evidence seems to make evidential support too easy to come by. In Bill Calculation, the view says that only after you acquire new evidence by performing a calculation does your total evidence support the conclusion that your shares are \$27.60. If this is right, then it seems hard to resist the view that S also acquires evidence between t_1 and t_2 , in this case *misleading* evidence. If in your case your total evidence at t_2 (consisting of the original evidence plus the evidence generated by your reasoning) supports the belief that your shares are \$27.60, then it seems that S 's evidence at t_2 (the original evidence plus the misleading evidence generated by S 's process of reasoning) would support the belief that your shares are \$27.10. This is a problematic result: we want to say that, at t_2 , S 's belief is not supported by S 's evidence and that your view is supported by your evidence. One might think that Reasoning as Evidence can acknowledge that reasoning can generate misleading evidence but still maintain that there is a crucial asymmetry in the evidential support you and S have for your respective beliefs. After all, despite S 's misleading evidence to the contrary, it seems that the original evidence still supports the \$27.60 rather than the \$27.10 answer. As we just saw, this response is not available: according to Reasoning as Evidence, the original evidence by itself supports neither the \$27.10 nor the \$27.60 answer. So the second problem is that Reasoning as Evidence is committed to the view that poor reasoning can lead to all sorts of beliefs having evidential support that intuitively are not supported by the subject's evidence.

Those who wish to defend conciliatory views on the basis of evidential considerations alone might be willing to swallow these counterintuitive consequences of Reasoning as Evidence.¹⁷ In that case, it still

17. One final consideration to mention at this point is that Reasoning as Evidence conflicts with the way some of the best defenders of conciliationism (see, for example, Christensen [2007, 2010b, 2011] and Feldman [2007a, 2007b]) set up the peer disagree-

strikes me as a significant cost to conciliationism that it would be committed to a very particular and controversial conception of evidential support. The problems associated with this conception warrant taking a serious look at alternative ways to defend conciliationism, which is what I will do in the next section.

5. Reasoning as Grounding

After section 3, we were left with the conclusion that conciliationism fails if we understand it in terms of ES, or similar standards of evidential support. But we were also left with a set of conciliatory intuitions about peer disagreement that need to be accounted for and a set of intuitions to the effect that, at t_1 , you would not be justified in believing that p . I just argued, in section 4, against the attempt to account for these intuitions in terms of a more expansive notion of evidential support. In this section, I propose to account for these intuitions and to defend conciliationism in terms of well-groundedness.

In Bill Calculation, it is intuitively plausible to think that, at t_1 , the belief that your shares are \$27.60 is supported by your evidence, but were you to believe that your shares are \$27.60, then this belief would not be well grounded in your evidence. This suggests the following principle:

Well-Groundedness: person S is justified in believing p at time t only if that belief is properly based on, or well grounded in, S 's evidence at t .

There is a great deal of debate about how to precisely characterize the well-groundedness or basing relation (see Korcz 1997 for an overview), but the important point for my purposes is that Well-Groundedness imposes a necessary condition on justified belief that is not captured by ES or by other standards of evidential support that make RR false.

ment problem. These authors are concerned with cases in which an epistemic peer disagrees with you given access to the same body of evidence. This doesn't happen according to Reasoning as Evidence: at t_1 you have the same evidence but no views about the issue yet, at t_2 and t_3 you have done some reasoning, hence generated some new evidence, so you no longer have access to the same body of evidence. It is not clear whether the arguments of these authors can be reformulated to fit this alternative framework. This is not to say that there is no interesting question of peer disagreement when we think about the issues in terms of evidential support alone, but it is unclear whether there is a strong case to be made for conciliationism when we do so.

In some cases, Well-Groundedness requires that the subject engages in the right kind of process of reasoning.¹⁸ Bill Calculation, intuitively, is a case like this. In contrast to Reasoning as Evidence, Well-Groundedness makes epistemic justification sensitive to certain nonevidential considerations. Because these considerations are facts about the belief and what it is grounded in, Well-Groundedness gives us a standard of what you are justified *in* believing, rather than what you are justified *to* believe.

I propose to think of conciliationism as a view about well-grounded belief. I will call this view "Reasoning as Grounding." According to this view, in peer disagreements generally, at t_3 , you are not justified in believing that p because that belief is not well grounded. In Bill Calculation, the view says that at t_2 , you have a well-grounded belief that your shares are \$27.60, but both at t_1 and t_3 your belief would fail to be well grounded. For t_1 it is clear why: you did not do any reasoning on the basis of your evidence about the bill. At t_3 , you have done the required reasoning, but here the conciliatory view says that the disagreement with your peer undermines the well-groundedness relation established by this reasoning. As a result, at t_3 you are in a situation that is in some respects similar to the one you are in at t_1 : you have sufficient evidence to support the belief that your shares are \$27.60, but without doing further reasoning you are not justified in believing that your shares are \$27.60.

This suggestion may seem implausible right off the bat. As I argue in section 3, at t_3 your evidence still supports the belief that your shares are \$27.60, despite your peer's disagreement. In addition, S 's disagreement does not change the past, so does not change how your belief was formed, and I assumed that at t_2 your belief that your shares are \$27.60 is well grounded. How could your peer's disagreement undermine the well-groundedness of your belief even though it leaves the belief supported by your evidence and leaves the causal history of your belief as it was? For certain very simple views of the well-groundedness relation, this question would have no answer, but more complicated conceptions of well-groundedness support my suggestion.

Here is one example of what such a more complicated conception might require. When some body of evidence E supports the belief that p , this is because there exists a certain relationship between E and p . In the

18. This is not true in all cases. For example, simple perceptual beliefs may be well grounded just when they stand in the right kind of causal relation to appearances.

case of inferentially justified beliefs,¹⁹ it is plausible to hold that this relationship has the structure of an argument: if E supports the inferentially justified belief that p, then there is at least one good argument from E to p. In the case of ES, a good argument for p given E is an argument that the subject would find sufficiently convincing on ideal reflection. Consider the following as a necessary condition for S's belief that p to be well grounded in S's evidence E: the argument on the basis of which S in fact believes p is or resembles a good argument from E to p.²⁰

This can explain how S's disagreement could undermine well-groundedness in Bill Calculation. At t_2 you have sufficient evidence, E, to believe that your shares are \$27.60, and you believe so on the basis of a good argument. At t_3 , you learn that your peer disagrees, and you now have E_{t_3} . As I argued, E_{t_3} supports the belief that your shares are \$27.60. But it may still be the case that the original argument, on the basis of which you believe that your shares are \$27.60, is no longer a good argument from the total evidence now available to you. After all, S's disagreement presents you with a potential defeater. Given this defeater, a good argument from E_{t_3} can no longer simply be a calculation from E to the conclusion that your shares are \$27.60; a good argument must also respond to S's disagreement as a potential defeater. Your evidence makes such a response available: you have sufficient evidence to conclude that S's disagreement is explained by a mistake on S's side. But the reasoning on which your belief is in fact based at t_3 is still your original process of reasoning, and this does not include such a response to this defeater. Therefore, the reasoning on which your belief, at t_3 , is based no longer is or resembles a good argument from your evidence to the conclusion that your shares are \$27.60. This would explain how your belief at t_3 is not well grounded in your evidence, even though at both t_2 and at t_3 , your belief is supported by your evidence, and even though S's disagreement does not change the causal history of your belief.²¹

19. Note that I already limited my discussion of peer disagreement cases to cases in which you and S both arrive at your beliefs after a process of reasoning. So the limitation to inferentially justified beliefs is no further restriction on the scope of my conciliatory view.

20. Both Foley (1987, 175–86) and Audi (1993, 233–73) discuss requirements along these lines in the context of more detailed and complete analyses of the well-groundedness relation. I think this kind of requirement could be built into a variety of different kinds of conceptions of the well-groundedness or basing relation (for an overview, see Korcz 1997).

21. One may worry that this view severs a plausible connection between evidential support and well-groundedness: if you have sufficient evidence to believe p and if you

This is a mere sketch of what a suitable conception of well-groundedness would be like, such that peer disagreement could undermine well-groundedness. This sketch suggests that it is not implausible to think of conciliatory views as views about well-groundedness, but it falls short of a defense of such views. A defense of Reasoning as Grounding needs to show that peer disagreement not only can undermine well-groundedness but that it does. Such an argument can be provided using Christensen's Integration and Independence principles, reformulated in terms of well-grounded belief.

Integration becomes the following:

Integration (WG): a subject is not justified in believing that p if: (a) he or she believes p , and he or she is justified in believing that some other subject, S , believes $\sim p$, (b) he or she is justified in believing that either his or her own or S 's mistake explains his or her disagreement with S about p , and (c) he or she is not justified in believing that S 's mistake, rather than his or her own mistake, explains his or her disagreement about p .

And Independence becomes this:

Independence (WG): when we determine what a subject is justified in believing about the explanation of his or her disagreement with S about p , we should bracket the subject's original reasoning about p .

These principles have an intuitive appeal similar to Christensen's original principles.

To get a conciliatory view about well-grounded belief, we need to establish (a), (b), and (c). The assumption that (a) is slightly stronger than before: we assume not just that you are justified to believe S disagrees with you but that you are justified in so believing. Similarly for the assump-

respond appropriately to that evidence and believe p , then, it seems, you are justified in believing p . On the proposed view, your evidence at t_3 supports the belief that your shares are \$27.60, you believe that your shares are \$27.60 after a correct process of reflection, yet your belief is not well grounded. The explanation is that the process of reflection you actually engaged in was sufficient for a well-grounded belief that your shares are \$27.60 given the evidence available to you at t_2 but not for a well-grounded belief given the evidence available to you at t_3 . This doesn't conflict with the plausible connection mentioned before, but we must keep in mind that an appropriate response to your evidence means an appropriate response to the evidence currently available to you, not the evidence available at some earlier point.

tion that (b): we have to assume that people have a justified background belief that their disagreements are explained by a mistake on either side. Strengthening these two assumptions does not seem especially problematic. So the central step is, once again, to use Independence (WG) and facts about the peer disagreement situation to show that (c): you are not justified in believing that S's mistake, rather than your own, explains your disagreement.²²

We obtain this result as follows. In Bill Calculation, at t_3 , your evidence relevant to the question of whose mistake explains your disagreement is the original evidence about your shares of the bill, evidence about the answers you and S arrived at after a quick mental calculation, and evidence about you and S being epistemic peers. Suppose that you believe, at t_3 , that S's mistake explains the disagreement. Would you be justified in believing this? This belief would have to be well grounded not just in the evidence about the answers at which you arrived and the evidence about you and S being epistemic peers. After all, that evidence by itself does not show that S, rather than you, made a mistake, so even the best process of reasoning on that evidence alone could not lead to a justified belief that S's mistake explains the disagreement. So if the belief that S's mistake explains the disagreement is to be well grounded, it must be because of a process of reasoning that draws on the original evidence about your shares of the bill. You engaged in such reasoning, and by assumption you did so correctly, between t_1 and t_2 . This is where Independence (WG) comes in: it requires us to bracket or disregard precisely this actual reasoning process between t_1 and t_2 . Other than this reasoning process, there is nothing that establishes the right kind of connection between your evidence about the bill and the belief that S's mistake explains your disagreement. As a result, this belief is not well grounded.²³

22. I will ignore two worries about Integration (WG) as it is presently formulated. The first is that one might avoid the challenge disagreement poses to your belief by simply not forming the belief that the person disagrees with you, or simply not having the belief that your disagreements are explained by a mistake on either side. The second is that given (c), simply not having the belief that S's mistake explains your disagreement is enough to undermine the well-groundedness of your belief that p. I think these worries can ultimately be answered by refining Integration (WG), but I will leave these refinements aside here.

23. We do not reach this conclusion in cases of nonpeer disagreement in which you have good reason to believe that your opponent is a less capable reasoner about the issue at hand. In these cases you can have a well-grounded belief that your opponent must have made the mistake that is unaffected by the independence principle: even aside from reasoning about the original evidence, you have good evidence and can believe on the

This gives us (c), which is what we needed for the conciliatory argument. So, suitably modified, Christensen's principles provide an argument for Reasoning as Grounding. The view suffers from none of the problems of the Reasoning as Evidence view.

I should note that as I formulate the view, conciliationism claims that at t_3 you are not justified in believing that p . Conciliationism is sometimes defended as the view that, at t_3 , you are justified in suspending judgment on p , and sometimes that you are justified, at t_3 , to assign a degree of confidence to p equal to the average between your own and S 's degrees of confidence at t_2 . Can my arguments be used to defend these stronger claims? On a minimal understanding of suspension of judgment as the absence of both belief and disbelief, my arguments would also support the conclusion that at t_3 you are justified in suspending judgment on p . If we understand suspension of judgment as a separate doxastic attitude falling under requirements of evidential support and well-groundedness, then we would need to provide some additional arguments to establish that at t_3 you are justified in suspending judgment on p . A further argument would also be needed to establish the so-called "equal weight view": my arguments do not show that, at t_3 , you would be justified in assigning a specific degree of confidence to p that is the average between the views of you and S at t_2 . I'm not sure whether these additional arguments can be provided, but it seems to me significantly less complicated to argue that you are justified in suspending judgment on p , even on a substantial understanding of this notion, than that you are justified in assigning some specific degree of confidence to p . In any case, the version of the conciliatory view I defend is weaker than these alternative views.

Reasoning as Grounding supports a range of judgments about Bill Calculation, and similar cases, that I take to be highly intuitive. The view itself says that both at t_1 and t_3 you are not justified in believing that your shares are \$27.60. At t_1 this is because you did not do any reasoning on the basis of your evidence at all; at t_3 because your process of reasoning is defeated or undermined by the disagreement with your peer. These judgments are fully consistent with what ES, and similar standards of evidential support, say about the case. So, we can hold at the same time that at t_1 , t_2 , and t_3 , the belief that your shares are \$27.60 is supported by your

basis of a good argument that your disagreement is explained by a mistake on your opponent's side.

evidence and that at all these times you are justified *to* believe that your shares are \$27.60. I think this is an intuitively plausible take on what happens in Bill Calculation. But Reasoning as Grounding also explains the lingering intuition that you are not justified in simply continuing to believe that your shares are \$27.60: that belief would not be well grounded in your evidence.²⁴

One final consideration in favor of Reasoning as Grounding is that it can answer an important objection Thomas Kelly (2010, 122–25) raises to conciliatory views. The objection is as follows: in Bill Calculation, at t_3 , you and S both have access to two kinds of evidence. First, the original evidence about the bill; second, the psychological evidence provided by your believing that your shares are \$27.60, S's believing that your shares are \$27.10, and the evidence for believing that S is your epistemic peer. Suppose that conciliationism says that you and S should, at t_3 , both be roughly equally confident in both the \$27.60 and the \$27.10 answers.²⁵ It seems that the psychological evidence provides equally strong support for both these answers. But you and S also have access to the original evidence about the bill, and that evidence strongly supports the \$27.60 answer and not the \$27.10 answer. So it seems that in favoring roughly equal confidence in both answers, conciliatory views are committed to the view that the original evidence doesn't matter: epistemic justification is sensitive to the psychological evidence only. According to Kelly (2010, 123), this would be an "extremely dubious consequence" of the conciliatory view.

Christensen's (2011, 4) strategy is to acknowledge the force of the objection but to argue that conciliationism can make room for the relevance of the original evidence: "it would, I think, be very damaging to Conciliationist views if their insistence on Independence amounted to insisting that one's original evidence was irrelevant to the rationality of the beliefs one ended up with after making one's conciliatory epistemic compromise. Nevertheless, it seems to me that, for two separate reasons, the Conciliationist position does not have this consequence."

24. Reasoning as Grounding also supports what Christensen and Feldman take to be the basic set-up of the peer disagreement problem. They take the basic case to be one of disagreement given access to the same body of evidence. We saw that this does not happen according to Reasoning as Evidence—after all, the reasoning process between t_1 and t_2 provided you and S with new evidence concerning p . In contrast, according to Reasoning as Grounding, the evidence available to you and S at t_3 can be the same.

25. Both Kelly and Christensen discuss these issues in terms of degrees of confidence. This difference will not be important for my argument here.

The first reason is that we can introduce a distinction between propositional and doxastic justification. The suggestion is that it is true that your own and S's postcompromise attitudes are equally justified in the propositional sense but that doxastic justification requires something like well-groundedness and that S's attitude would fail to be well grounded because S's original reasoning was faulty (see Christensen 2011, 4). Hence your original reasoning and evidence are not irrelevant for the justification of your beliefs. I agree that an appeal to considerations of well-groundedness (what Christensen calls "doxastic justification") can answer Kelly's objection. But it cannot be done in the way Christensen proposes. On his view, conciliationism is correct even if we think about the issue exclusively in terms of propositional justification, even if we don't take well-groundedness into consideration. Well-groundedness only comes in when we consider the additional demands of doxastic justification. I argued that this is false: we need to appeal to considerations of well-groundedness to arrive at a conciliatory view in the first place.

The second reason is that Christensen thinks that conciliationism specifies what the appropriate response is to one particular piece of evidence—the evidence provided by the peer disagreement—not what one is justified to believe, all things considered. In the simple calculation case, Christensen thinks that *you* would be justified to be roughly equally confident in the \$27.60 and the \$27.10 answers, all things considered. But this is not true for S. For S, the evidence provided by the disagreement requires S to move in the direction of the \$27.60 answer, but for S the original evidence provides even further reason to become confident that the answer is \$27.60.²⁶ If both of you are, at t_3 , roughly equally confident in the competing answers, then your belief is justified while S's belief is not. As a result, the original evidence is relevant for the epistemic justification of the beliefs of the disagreeing parties, even on a conciliatory view.

I think this response to Kelly's objection is problematic for a number of reasons. I will state only one. The problem is that Kelly's objection comes up in a different place even if we accept everything Christensen

26. Christensen's view here is complicated. He says that there is a sense in which S should correctly assess the original evidence, correct his or her mistake, and become highly confident that your shares are \$27.60. But Christensen thinks that the peer disagreement case supposes that S does not correct her original mistake. Holding fixed this mistake, Christensen (2011, 4) thinks there is nothing S can do such that her belief is fully rational.

says in response to it. As we just saw, Christensen thinks that conciliationism is committed to the view that in Bill Calculation, *you* are justified to be roughly equally confident in the competing answers, all things considered. Imagine an alternative version of Bill Calculation in which you have lost, at t_3 , your original evidence. In particular, you forgot what the total on the bill is. In that case too, it seems that at t_3 , you should be roughly equally confident in the competing answers. If anything, it seems that this conclusion is even more plausible than in the original case. But now we have Kelly's problem back. In the original Bill Calculation case and in the forgotten evidence variant, you are, at t_3 , in epistemically identical situations, except that in the original case you have access to a piece of evidence (that the total is forty-six dollars) that strongly supports one of the competing answers (that your shares are \$27.60). Yet, the conciliatory view would have it that in both cases you should have roughly the same degree of confidence in both answers. Once again, having access to the original evidence does not make any difference.

The general lesson is that it is hard to acknowledge the force of Kelly's objection and hold on to a conciliatory view. Once we acknowledge that the original evidence plays an important role in determining what *S* is justified to believe, it becomes less clear why the availability of the original evidence wouldn't also skew what *you* are justified to believe in the direction of the \$27.60 answer. When we understand conciliationism as a view about well-grounded belief, we can respond to Kelly's objection much more straightforwardly: we simply do not acknowledge the force of the objection. When justified belief is understood to be well-grounded belief, it is not at all uncommon or surprising that the availability of a significant body of evidence relevant to *p* makes no difference to the belief about *p* that you would be justified in having. In fact, Bill Calculation provides us with an example of this. At t_1 but not before that (at t_0), you have evidence that strongly supports the belief that your shares are \$27.60. Yet at both these times, the belief that your shares are \$27.60 would be unjustified. Until you go through the appropriate process of reasoning, the evidence available to you at t_1 makes no difference for what you are justified in believing, despite the fact that the evidence strongly supports the belief that your shares are \$27.60. The fact, pointed out by Kelly, that a situation like this arises at t_3 as well is no objection to conciliatory views.

This is not to accept Kelly's charge that conciliationism makes the original evidence altogether irrelevant. Conciliationism is merely com-

mitted to the view that in certain cases both your belief that p and S 's belief that $\sim p$ are not well grounded, even though you both have a body of evidence available to you that supports the belief that p .

6. Concluding Remarks: Consequences for the Peer Disagreement Debate

The main argument of this article has been as follows: conciliatory views are false if we understand them to be views about any conception of evidential support out of a wide range of such conceptions. We could attempt to save the conciliatory argument by appeal to some conception of evidential support that falls outside this range, but I argued that the prospects of this strategy look bleak. So far, this is bad news for conciliationism: if we look at the peer disagreement debate purely in evidentialist terms, conciliationism will not succeed. I then argued for some good news for conciliationism: conciliationism can be saved if understood as a view about well-grounded belief. I argued that Christensen's argument for conciliationism can be made to work if we appeal to considerations of well-groundedness, that the resulting view leads to an intuitively correct picture of what happens in cases like Bill Calculation, and that we can answer an important objection raised by Kelly.

By themselves, these are interesting conclusions, but I end this article with some remarks about the broader significance of these arguments for the peer disagreement debate. First, my discussion has some important methodological consequences. One of the most prominent argumentative strategies in the peer disagreement debate is as follows. One constructs a particular example of a peer disagreement. One generates an intuition about what the parties should believe, have reason to believe, are justified to believe, or would be rational in believing, about the disputed proposition. One uses this intuition to defend or attack a certain view about the epistemic significance of peer disagreement.²⁷ This strategy is flawed. My discussion shows that the central questions raised about these cases are ambiguous between different standards of epistemic justification: when we ask what one should believe, or what it would be rational to believe, we cannot assume that we have specified sufficiently what we are asking about.

27. This strategy is found throughout the debate; see Christensen 2007, 2011; Kelly 2005, 2010; Lackey 2010; Feldman 2007a; and many of the contributions in Feldman and Warfield 2010.

As a result, it is unclear what kind of epistemic standard the intuitions generated in this way track. Bill Calculation provides a good example of this: at t_1 there is a perfectly legitimate intuition that you should believe that your shares are \$27.60, because that is the view that your evidence supports, but there is also a perfectly legitimate intuition that you should not believe that your shares are \$27.60, after all, you did not do any reasoning about the issue. My discussion shows that this kind of conflict of intuitions also occurs in cases of peer disagreement. In fact, my discussion suggests, though I do not argue for this claim, that the debate between two of the most prominent participants in the debate, Thomas Kelly and David Christensen, turns on precisely this kind of ambiguity. Of course, this is not a general argument against the appeal to intuitions about epistemic justification. The point is that we cannot simply appeal to intuitions about “what one should believe” without further specification.

Finally, my discussion suggests that the current debate leaves untouched a number of interesting questions about the epistemic significance of peer disagreement. Most of the debate consists of defenses of, and attacks on, conciliationism. My arguments suggest that conciliationism is a view about one particular aspect of the epistemic significance of peer disagreement: it’s a view about well-grounded belief. But there is a separate and independently interesting question about the impact of peer disagreements on evidential support. I argued that in cases like Bill Calculation, the evidence provided by your disagreeing peer is not very weighty, but this may be different in other cases of peer disagreement. In fact, I think that whereas conciliationism may show us what peer disagreement cases have in common—they undermine well-groundedness—a discussion of the evidential impact of peer disagreement may help us understand interesting distinctions between different cases of peer disagreement.

References

- Audi, Robert. 1993. *The Structure of Justification*. Cambridge: Cambridge University Press.
- Ballantyne, Nathan, and E. J. Coffman. 2012. “Conciliationism and Uniqueness.” *Australasian Journal of Philosophy* 90: 657–70.
- Christensen, David. 2007. “Epistemology of Disagreement: The Good News.” *Philosophical Review* 116, no. 2: 187–217.
- . 2009. “Disagreement as Evidence: The Epistemology of Controversy.” *Philosophical Compass* 4, no. 5: 756–67.

- . 2010a. "Rational Reflection." *Philosophical Perspectives* 24, no. 1: 121–40.
- . 2010b. "Higher-Order Evidence." *Philosophy and Phenomenological Research* 81, no. 1: 185–15.
- . 2011. "Disagreement, Question-Begging and Epistemic Self-Criticism." *Philosophers' Imprint* 11, no. 6: 1–22.
- Elga, Adam. 2007. "Reflection and Disagreement." *Noûs* 41, no. 3: 478–502.
- Enoch, David. 2010. "Not Just a Truthometer: Taking Oneself Seriously (But Not Too Seriously) in Cases of Peer Disagreement." *Mind* 119, no. 476: 953–97.
- Feldman, Richard. 2007a. "Epistemological Puzzles about Disagreement." In *Epistemology Futures*, ed. Stephen Hetherington, 216–36. Oxford: Oxford University Press.
- . 2007b. "Reasonable Religious Disagreement." In *Philosophers without Gods*, ed. Louise Antony, 194–214. Oxford: Oxford University Press.
- . 2009. "Evidentialism, Higher-Order Evidence, and Disagreement." *Episteme* 6: 294–312.
- Feldman, Richard, and Earl Conee. 2008. "Internalism Defended." In *Epistemology: An Anthology*, 2nd ed., ed. Ernest Sosa, Jaegwon Kim, Jeremy Fantl, and Matthew McGrath, 407–21. Oxford: Blackwell.
- Feldman, Richard, and Ted A. Warfield, eds. 2010. *Disagreement*. Oxford: Oxford University Press.
- Foley, Richard. 1987. *The Theory of Epistemic Rationality*. Cambridge, MA: Harvard University Press.
- . 1993. *Working without a Net: A Study of Egocentric Epistemology*. Oxford: Oxford University Press.
- Goldman, Alvin. 2010. "Epistemic Relativism and Reasonable Disagreement." In Feldman and Warfield 2010, 187–215.
- Gutting, Gary. 1982. *Religious Belief and Religious Skepticism*. Notre Dame, IN: University of Notre Dame Press.
- Kelly, Thomas. 2005. "The Epistemic Significance of Disagreement." In *Oxford Studies in Epistemology*, Vol. 1, ed. Tamar Gendler Szabó and John Hawthorne, 167–96. Oxford: Oxford University Press.
- . 2010. "Peer Disagreement and Higher-Order Evidence." In Feldman and Warfield 2010, 111–74.
- Korcz, Keith. 1997. "Recent Work on the Basing Relation." *American Philosophical Quarterly* 34, no. 2: 171–91.
- Kornblith, Hilary. 2010. "Belief in the Face of Controversy." In Feldman and Warfield 2010, 29–52.
- Lackey, Jennifer. 2010. "A Justificationist View of Disagreement's Epistemic Significance." In *Social Epistemology*, ed. Adrian Haddock, Alan Millar, and Duncan Pritchard, 298–325. Oxford: Oxford University Press.
- Lam, Barry. 2011. "On the Rationality of Belief-Invariance in Light of Peer Disagreement." *Philosophical Review* 120, no. 2: 207–45.

- Matheson, Jonathan. 2009. "Conciliatory Views of Disagreement and Higher-Order Evidence." *Episteme* 6: 269–79.
- White, Roger. 2005. "Epistemic Permissiveness." *Philosophical Perspectives* 19: 445–59.
- van Wietmarschen, Han. n.d. "The Epistemic Significance of Deep Disagreement." Unpublished manuscript.

Truth, the Liar, and Relativism

Kevin Scharp

The Ohio State University

1. Introduction

I aim to establish a new connection between two topics. The first is the *aletheic paradoxes* (that is, the paradoxes affecting truth, of which the liar is merely the most famous).¹ Nearly as old as Western philosophy itself, work on the aletheic paradoxes is still vibrant today. Contributions to this topic from analytic philosophy have come in roughly three waves. The first wave is based on Alfred Tarski's work from the 1930s, which gave truth conditions for formulas of first-order predicate calculus and set the stage for much of what came after. Saul Kripke's seminal paper from 1975 posed serious problems for applying Tarski's results to natural language and used new mathematical techniques to address the paradoxes. Kripke's work also inspired a whole generation of new approaches in the 1980s and early 1990s.² After something of a lull, the twenty-first century has seen a flurry of new activity—enough to constitute a third wave that is still building.³

Portions of this essay have been read at the University of St Andrews, the University of Richmond, the University of North Carolina at Chapel Hill, University of Melbourne, The Ohio State University, University of Aberdeen, and Ludwig Maximilians University; many thanks to those audiences for their help in improving these ideas. I also thank Stewart Shapiro, Neil Tennant, Matti Eklund, David Ripley, Raleigh Miller, Judith Tonhauser, and several anonymous referees for extensive comments on earlier drafts.

1. I use 'aletheic' to mean *pertaining to truth*.

2. These include Charles Chihara (1979), Graham Priest (1979), Tyler Burge (1979), Anil Gupta (1982), Stephen Yablo (1985), Jon Barwise and John Etchemendy (1987), Vann McGee (1991), and Keith Simmons (1993).

3. Third-wave theorists include Matti Eklund (2002a), Michael Glanzberg (2004),

The second topic is old as well—relativism. Probably around the time Eubulides allegedly formulated the liar paradox (circa fourth or fifth century BCE), Protagoras gave a voice to the view that man is the measure of all things. Since then, formulations of relativism have become more precise but have not fundamentally changed until this century, just a few years ago, when relativism was cast as a semantic theory, or rather, as a family of presemantic, semantic, and postsemantic theories (these terms are explained below); nonindexical contextualism, judge-dependence, and assessment-sensitivity are the most familiar. The debate over their merit has just begun but has accelerated to a breakneck pace. Many of the people involved approach the topic with their own idiosyncratic terms and classifications, which makes the literature difficult to follow. Such is a philosophical topic at its inception. There are few maps of this shifty terrain, but most agree that a semantic relativist about some topic takes truth, insofar as it pertains to that topic, to be relative to some unorthodox parameters.

To my knowledge, no analytic philosophers have connected these two ancient topics. No one I know of says that relativism is an effect of the alethic paradoxes or serves as an antidote to them. Nor does anyone diagnose the alethic paradoxes as caused by relativism or link solutions to them as a way of explaining away or dealing with the consequences of relativism. So, the connection forged here is novel but perhaps not totally unexpected, given the prominent role truth plays in each topic.

I propose a solution to the alethic paradoxes on which truth predicates are assessment sensitive. Truth is not an antecedently plausible topic for a semantic relativist treatment; nevertheless, the alethic paradoxes give us good reason to think that truth is an inconsistent concept, and there are good reasons to think that semantic relativism is appropriate for inconsistent concepts. Thus, I show that a promising version of the best approach to the paradoxes is an application of semantic relativism to truth itself—arguing from results about the paradoxes and general considerations about language use to alethic assessment-sensitivity.

The essay is divided into two parts, the first on the alethic paradoxes and the second on assessment-sensitivity with respect to truth predicates. The first contains an overview of my preferred approach to the paradoxes, which entails that truth is an inconsistent concept and should

Tim Maudlin (2004), Lionel Shapiro (2006), Hartry Field (2008), Jc Beall (2009), Volker Halbach (2011), Leon Horsten (2011), Elia Zardini (2011), and David Ripley (2012).

be replaced (for certain purposes) by a team of consistent concepts that can do its work without causing troubling paradoxes. The second part considers which treatment is most appropriate for our inconsistent concept of truth. In it, I propose an assessment-sensitivity view of truth, reply to several objections, and review some issues that arise for approaches to the alethic paradoxes.

2. The Alethic Paradoxes

The liar paradox is the best-known symptom of a major problem with our reasoning about truth. It pertains to sentences like:

- (1) (1) is not true.

Although there are many ways the argument might go, it is most common to appeal to the following principles about truth:

- (T-In) If p , then $\langle p \rangle$ is true.
(T-Out) If $\langle p \rangle$ is true, then p .
(Sub) If $s = t$, and $\langle p \rangle$ results from replacing some occurrences of ' s ' with ' t ' in $\langle q \rangle$, then $\langle p \rangle$ is true iff $\langle q \rangle$ is true.

In these principles, ' p ' and ' q ' are sentential variables—they are placeholders for sentences—and ' $\langle p \rangle$ ' and ' $\langle q \rangle$ ' are quote-names for the sentences that fill in for ' p ' and ' q ', respectively; ' s ' and ' t ' are singular-term variables. From these principles about truth and intuitive logical principles, one can derive that (1) is true and (1) is not true.⁴

There have been, to put it mildly, many suggestions for how to deal with the alethic paradoxes. The vast majority of them are *traditional* approaches in the sense that they single out a step (or steps) in the reasoning as fallacious. The fallacious move is taken to be either:

- (i) a principle about truth (for example, (T-In)),⁵
(ii) a logical principle (for example, excluded middle),⁶ or

4. Two other familiar paradoxes affect truth. The first is Curry's paradox: if $p =$ 'if p is true, then $0 = 1$ ', then one can derive ' $0 = 1$ ' (or any other absurdity); see Curry 1942. The second is Yablo's paradox: if $p_n =$ 'for all $m > n$, p_m is false' for $n \geq 0$, then one can derive that each p_n is both true and false; see Yablo 1993c. As I use the term, all three are alethic paradoxes. There are many others not fortunate enough to have names; see figure 1.

5. See Maudlin 2004 for an example and Field 2008, chaps. 6–14, for a survey.

6. See Field 2008, chaps. 15–23, for an example.

- (iii) an assumption about language or thought (for example, sentence (1) is meaningful).⁷

A traditional approach to the paradoxes is typically accompanied by two major worries. First, although it avoids the liar paradox and the other familiar alethic paradoxes, it encounters paradoxes that are structurally similar to these but involve truth and some other notions—these have come to be called *revenge paradoxes*. The worry is that if a traditional approach deals with the familiar alethic paradoxes but faces revenge paradoxes, then it does not qualify as an acceptable approach.⁸ Second, the principles involved in the derivation of the alethic paradoxes are all so obvious that they seem to be constitutive of the concepts involved. That is, it seems like rejecting one of the principles involved merely changes the subject rather than solves the paradoxes. The worry is that it simply does not seem plausible that we could be mistaken in the way indicated by a traditional approach; so traditional approaches have a major credibility problem when it comes to their diagnoses.

There is an alternative to traditional approaches that seems to avoid the second worry—paraconsistent dialetheism.⁹ According to the dialetheist, the liar reasoning is sound, which implies that the conclusion, ‘(1) is true and (1) is not true’ is true. In order to avoid triviality (that is, a trivial consequence relation where every formula is a consequence of every set of formulas), the dialetheist endorses a nonclassical logic in which the rule *ex falso quodlibet*—aka explosion—(that is, $p, \sim p \vdash q$) is invalid, and some contradictions are true. At best, dialetheism requires a major revision in what we take to be logically valid inferences, but there are a host of other costs as well.¹⁰ However, it seems to me that the biggest worry is that, to many, the very idea that some contradictions are true is unintelligible.

Notice that, despite their differences, we can think of dialetheism as just another traditional approach that rejects a logical principle, except in this case it is explosion (and the other inference rules linked to it) that we are asked to give up. Instead of making a mistake in the liar reasoning itself, the dialetheist thinks our mistake lies in rejecting its conclusion.

7. See Sorensen 2001 for an example.

8. Although I touch on revenge paradoxes in what follows, they are not the focus of this work—see Shapiro 2011; Scharp 2013, forthcoming; and the papers in Beall 2008 for discussion.

9. See Priest 2006a, 2006b; and Beall 2009.

10. See papers in Priest, Beall, and Armour-Garb 2005 for discussion.

The common core of traditional approaches and dialethic approaches is the assumption that we can simply reject one of the principles involved in the liar paradox. Both types of approaches invite a similar worry—the principles involved in the derivation of the paradox seem to be *constitutive* of the concepts involved; (T-In) and (T-Out) seem to be constitutive of the concept of truth, and the logical principles seem to be constitutive of the logical concepts used in the reasoning. If that is right, then it seems that we cannot give these principles up without also giving up the concepts in question along with them. That is, traditional and dialethic approaches seem to “solve” the alethic paradoxes by telling us to stop using the concepts involved. Granted, there are many questions surrounding the nature of concepts and constitutive principles, but these should not distract one from the underlying worry—it seems like our very competence with the concepts involved leads us to accept the assumptions and the inferences in the liar reasoning and to reject its conclusion. If that is correct, then traditional approaches and dialethic approaches to the alethic paradoxes are deeply flawed.¹¹

It seems to me that any plausible approach to the alethic paradoxes must explain why we are so taken in by the paradoxical reasoning—why we find it so difficult to accept any of the traditional approaches or dialetheism. There is a small but growing tradition of philosophers who accept this condition as well; although there is no consensus on what to call it yet, I use the term ‘the inconsistency approach’, for reasons that will become clear shortly.

According to inconsistency approaches, we are taken in by the liar paradox by virtue of our conceptual or linguistic competence. If that is correct, then there is something seriously wrong with at least some of the concepts involved. Inconsistency theorists agree that the obvious culprit here is the concept of truth.¹² Indeed, on the formulation I prefer, we can say that our concept of truth is inconsistent. Others in this tradition prefer to say that a *language* that contains a truth predicate is inconsistent.¹³ I do not see that anything hangs on this difference.

11. See Eklund 2002b, 2008b; and Beall and Priest 2007 for a discussion of some of these points.

12. One might think that one or more of our logical concepts should be blamed instead. Having argued against this point elsewhere, I am not going to rehash the case for blaming truth instead of our logical notions. See Scharp 2008, 2013, forthcoming.

13. See, for example, Chihara (1979, 1984), Eklund (2002a, 2002b, 2007, 2008a), and Patterson (2006, 2007, 2009).

Two prominent dialetheists, Jc Beall and Graham Priest, strike back by arguing that the most well-known inconsistency view around, Matti Eklund's theory, does not provide the most essential element in an approach to the alethic paradoxes; namely, an account of truth. They write:

Eklund... does not even say what his own account of truth is.... All standard accounts of truth, including Kripke's, run into problems of well known kinds (which, again, have nothing whatsoever to do with analyticity, and so to which Eklund's distinction is irrelevant). All are subject, for example, to strengthened versions of the liar paradox. Thus, if one takes it that the liar sentence is neither true nor false (as does Kripke), then one has only to consider the sentence 'This sentence is not true'. If it is neither true nor false, it is not true, and so true. Moreover, there are reasons as to why problems of this kind would seem to be inevitable. Such things are well known and there is no need to dwell on them here. We note them only to point out that what is needed to solve the liar paradox is an account of truth, not of the meaning-theoretic status of its principles. Since Eklund does not engage with this issue, there is, in a sense, nothing in his paper to answer. (Beall and Priest 2007, 79)

Beall and Priest (*ibid.*) conclude that "Eklund's failure to spell out an account of truth" leaves "him 'out of the game' (as it were)." Eklund (2008b, 100) replies to their charge of being "out of the game" by writing, "They are right that there is a particular debate on the liar—indeed, the most central debate—where I do not explicitly have a stand. I do not have, in the terminology earlier introduced, a positive theory of the liar." I imagine that Beall and Priest would say the same about most of the other members of the inconsistency tradition (for example, Chihara [1979], Tappenden [1993], Patterson [2006], Burgess and Burgess [2011], and me [2007, 2008]).¹⁴ For what it is worth, I agree with them. The central purpose of this essay is to get in the game—to provide a theory of truth in the sense of the above quotation that is the centerpiece of an inconsistency approach to the alethic paradoxes. That is, the aim is to provide a consistent theory of our inconsistent concept of truth that can compete with the other players.¹⁵

14. Note that Yablo is an inconsistency theorist who is in the game, but his views are often ignored for some reason—see Yablo (1985, 1993a, 1993b).

15. At the end of this essay, I provide a categorization scheme for approaches to the alethic paradoxes—in the terminology introduced there, being in the game requires one to have a logical approach to the paradoxes.

I am not going to argue that truth is an inconsistent concept or that an inconsistency approach is superior to the other kinds of approaches to the alethic paradoxes; I have done so in other work (see Scharp 2008, 2013). Nor am I going to argue that the kind of inconsistency theory I endorse is better than the other members of this tradition; this has been done elsewhere as well (see Scharp 2007, 2008). Instead, the focus of this essay is on the details of what I take to be the right kind of inconsistency approach to the liar.

The central claim of the approach I offer, and what distinguishes it from its rivals in the inconsistency tradition, is that we need to replace the concept of truth with a team of concepts. These replacements can do the work we require of truth, but they are free of its defect (that is, they do not give rise to paradoxes). In addition, they can be used to give an illuminating explanation of our concept of truth, which is the focus of this essay.

2.1. Inconsistent Concepts

The approach I advocate can be summed up in the claim that truth is an inconsistent concept. Of course, that is not very informative unless one knows what inconsistent concepts are. In this subsection, I introduce the idea.

As a rough first step, consider the following definition: a concept is inconsistent if and only if by virtue of one's competence with it, one is led to accept inconsistent beliefs. Note that one is not *inevitably* led to accept inconsistent beliefs, but one's conceptual competence in these cases can make some inconsistent beliefs *seem* inevitable.

It would be nice to have an example. What I see as inconsistent concepts have figured in philosophical discussions outside of the literature on truth—Arthur Prior (1960) introduced the connective 'tonk' in an attempt to undermine inferential-role theories of the logical operators, Michael Dummett (1973) discussed the pejorative 'Boche' while motivating his views on meaning and assertibility conditions, and Hartry Field (1973) focused on 'mass' as it was used in Newtonian mechanics in his papers on referential indeterminacy.¹⁶ I use Field's 'mass' example as an analogy.

In Newtonian mechanics, physical objects have a physical quantity, mass. According to this theory, mass obeys two laws (which are considered equally fundamental): (i) $\text{mass} = \text{momentum} / \text{velocity}$, and (ii) the mass

16. See also Gupta 1999.

of an object is the same in all reference frames. In relativistic mechanics, physical objects have two different “kinds” of mass: proper mass and relativistic mass. An object’s proper mass is its total energy divided by the square of the speed of light, while an object’s relativistic mass is its nonkinetic energy divided by the square of the speed of light. Although relativistic mass = momentum/velocity, the relativistic mass of an object is not the same in all reference frames. On the other hand, proper mass $\neq$ momentum/velocity, but the proper mass of an object is the same in all reference frames. Thus, relativistic mass obeys one of the principles for mass, and proper mass obeys the other. Since we live in a relativistic universe (that is, one where momentum over velocity is not the same in all reference frames), mass is an inconsistent concept. That is, before the twentieth century, we used a concept whose constitutive principles are inconsistent with what would come to be certain well-confirmed claims about the world (for example, that momentum/velocity is not the same in all reference frames).¹⁷

I want to point out four important features of the mass example. First, if the natural laws of our universe had been different, then the concept of mass might not have been an inconsistent concept. That is a common feature of inconsistent concepts—they are inconsistent relative to an environment in which they are used. Second, no amount of introspection, conceptual analysis, or reflection would have allowed possessors of the concept of mass in the nineteenth century to determine that it is inconsistent. Because it is inconsistent relative to the natural laws of our world, finding out that the concept of mass is inconsistent is an empirical discovery. Thus, at least for some inconsistent concepts, simply possessing them is not sufficient for being in a position to know that they are inconsistent. Third, it would not have made sense to avoid the inconsistency described above by giving up one or more logical principles involved in the reasoning in question. The problem was obviously with Newtonian mechanics and its concept of mass. Fourth, inconsistent concepts can be useful. The concept of mass was extraordinarily useful and did not cause its possessors any trouble for centuries. Moreover, we still use it

17. Field does not say that the concept of mass is inconsistent; that is my formulation. However, he does say that ‘mass’ figures in two “central tenets of Newtonian mechanics” that are “both extremely central to Newton’s theorizing and to his scientific practice.” Field 1973, 101, 102. If these central tenets are constitutive of the concept of mass, then mass is an inconsistent concept. For discussion of this example, see Earman and Fine 1977, Field 1994, Jammer 2000, and Petkov 2009.

frequently. If someone wants to design a house, the concept of mass will work just fine. The challenge is to understand this practice—how can we still use an inconsistent concept knowing full well that it is inconsistent?

With these features in mind, I want to present a better definition of ‘inconsistent concept’:

- (IC) A concept χ is *inconsistent* if and only if $\Gamma \cup \Phi$ is an inconsistent set, where Γ is the set of χ ’s constitutive principles, and Φ is a set of facts.

Given that momentum/velocity is not the same in all reference frames, the concept of mass is inconsistent. My aim in this essay is to sketch a theory of truth on which truth is an inconsistent concept.

2.2. *Descriptive Projects and Prescriptive Projects*

In the late 1970s and early 1980s, several theorists working on the liar paradox reflected on what one might be doing in presenting an approach to it. Following Charles Chihara, Anil Gupta, and Stephen Yablo, we can identify the following projects:

Diagnostic project: specifies and explains the causes of the alethic paradoxes.

Preventative project: says what must be done to keep them from arising in artificial languages that can be used to model our semantic concepts.

Descriptive project: explains the use and meanings of truth predicates in natural language.

- (i) *Semantic project*: provides a philosophical theory that systematically yields semantic results that accord with the intuitions of natural language speakers.
- (ii) *Psychological project*: specifies how natural language speakers arrive at their views on semantic issues (for example, the truth values of sentences).

Prescriptive project: specifies how we should change our natural language in light of the alethic paradoxes.¹⁸

18. Chihara (1979, 590–91) discusses the diagnostic project, the preventative project, and the prescriptive project. Gupta (1982, 1–2) distinguishes the descriptive project from the prescriptive project. Yablo (1985, 297–300) distinguishes the semantic project from the psychological project.

My aim in this essay is to provide the details of the descriptive project (semantic version). Along the way, I address the diagnostic project and the prescriptive project as well; indeed, this entire section is dedicated to them. I have already mentioned my view on the diagnostic project—the alethic paradoxes arise because truth is an inconsistent concept. In the following subsections, I say more about my view on the prescriptive project, but I have already indicated my commitment to replace our inconsistent concept of truth because it is imperative that *the descriptive project does not employ the concept of truth*.¹⁹ Instead, if the descriptive project offers the kind of theory that traditionally employs the concept of truth, it should be altered so as to employ the replacement concepts instead. In trying to specify the semantic features of the truth predicate, especially when it runs into trouble, we should not be using the truth predicate. The upshot is that the descriptive theory I offer appeals to the replacement concepts posited by the prescriptive theory.

2.3. *The Prescriptive Project*

Given that truth is an inconsistent concept, how should we change our natural language or conceptual scheme? First off, changing our logic is not a good option for the inconsistency theorist, who sees the alethic paradoxes as symptoms of a defect in our concept of truth, not our logical concepts. Although there is a place for disputes about which logic (or logics) people use (or should use), none of the standard alternatives (for example, intuitionistic logic (I), the logic of entailment (E), and the logic of relevant implication (R)) are weak enough to avoid the alethic paradoxes. So it seems that considerations about logic or reasoning alone will not solve them. Moreover, since the inconsistency theorist thinks that our concept of truth has constitutive principles that are inconsistent (relative to some claims about the world) and does not accept contradictions, he or she will not accept that all of truth's constitutive principles are true. Thus, the inconsistency theorist's position is that we should reject one or more of truth's constitutive principles that lead to the contradictions via reasoning in the alethic paradoxes. Finding fault with one of the logical principles involved in the reasoning as well would be overkill. In addition,

19. I argue in Scharp 2007 that if Eklund's inconsistency theory of truth is correct, then the concept of truth to which it appeals cannot do the work required of it by the theory; similar problems confront any descriptive theory of our inconsistent concept of truth that casts the concept of truth in an explanatory role.

from the perspective of an inconsistency theorist, our response to the defect in our concept of truth should be no different than our responses to defects in other useful concepts throughout history. We did not try to keep Newtonian mechanics and its concept of mass in light of conflicting empirical discoveries by revising our logic to avoid inconsistencies in our belief system. Instead, we admitted that the Newtonian theory and the concepts it implicitly defines need to be replaced for certain purposes.²⁰ Of course, an inconsistency theorist might not accept classical logic, but these considerations suggest that as a methodological principle, an inconsistency theory of truth should be compatible with classical logic so that it is as uncontroversial as possible.

An inconsistency theorist who thinks we should replace our concept of truth (at least for certain purposes) has a choice to make: what should the replacement(s) be? It is tempting to opt for a single replacement concept, but there are good reasons to reject this strategy.

First, it is widely accepted that we use 'true' to endorse propositions that we cannot assert directly; for example, Ralph can assert 'the Riemann hypothesis is true' and thereby endorse the Riemann hypothesis even though he does not remember or has never learned which sentence expresses it, or he can assert 'all the axioms of ZFC are true' and thereby endorse all the axioms of ZFC even though there are too many for him to assert one by one. We can capture this role by saying that a truth predicate functions as *a device of endorsement*. The flip side of this role is *a device of rejection*; he can say 'the Riemann hypothesis is not true' and thereby reject the Riemann hypothesis. In order to serve as a device of endorsement, the truth predicate must obey (T-Out), and in order to serve as a device of rejection, the truth predicate must obey (T-In). Of course, we already know that in a classical setting, no single concept obeys these two principles; thus, no concept can serve as both a device of endorsement and rejection given classical logic and the expressive resources to construct liar sentences. However, as I shall show, if we replace truth with two concepts, we can split the workload, allowing one to serve as a device of endorsement and the other to serve as a device of rejection.

The huge variety of alethic paradoxes hidden in the principles that truth seems to obey constitutes a second reason to use a team of replacements. Here is a sample of the many principles that truth seems to obey:

20. See McGee 1991, 102–3, for a similar argument and Horsten 2011, 136–7, for an alternative view.

Disquotational Principles

(T-Out)	$T(\langle p \rangle) \rightarrow p$
(T-In)	$p \rightarrow T(\langle p \rangle)$
(T-Elim)	$T(\langle p \rangle) \vdash p$
(T-Intro)	$p \vdash T(\langle p \rangle)$
($\sim$ T-Elim)	$\sim T(\langle p \rangle) \vdash \sim p$
($\sim$ T-Intro)	$\sim p \vdash \sim T(\langle p \rangle)$
(Cat)	$\vdash p \rightarrow \vdash T(\langle p \rangle)$
(Co-Cat)	$\vdash T(\langle p \rangle) \rightarrow \vdash p$

Truth-Functional Principles

($\sim$ -Imb) ²¹	$\sim T(\langle p \rangle) \rightarrow T(\langle \sim p \rangle)$
($\sim$ -Exc) ²²	$T(\langle \sim p \rangle) \rightarrow \sim T(\langle p \rangle)$
($\wedge$ -Imb)	$T(\langle p \rangle) \wedge T(\langle q \rangle) \rightarrow T(\langle p \wedge q \rangle)$
($\wedge$ -Exc)	$T(\langle p \wedge q \rangle) \rightarrow T(\langle p \rangle) \wedge T(\langle q \rangle)$
($\vee$ -Imb)	$T(\langle p \rangle) \vee T(\langle q \rangle) \rightarrow T(\langle p \vee q \rangle)$
($\vee$ -Exc)	$T(\langle p \vee q \rangle) \rightarrow T(\langle p \rangle) \vee T(\langle q \rangle)$
($\rightarrow$ -Imb)	$(T(\langle p \rangle) \rightarrow T(\langle q \rangle)) \rightarrow$ $T(\langle p \rightarrow q \rangle)$
($\rightarrow$ -Exc)	$T(\langle p \rightarrow q \rangle) \rightarrow (T(\langle p \rangle) \rightarrow$ $T(\langle q \rangle))$

Miscellaneous Principles

(Taut)	$T(\langle p \rangle)$ for p a tautology
(Contra)	$\sim T(\langle p \rangle)$ for p a contradiction
(T-Del)	$T(\langle T(\langle p \rangle) \rangle) \rightarrow T(\langle p \rangle)$
(T-Rep)	$T(\langle p \rangle) \rightarrow T(\langle T(\langle p \rangle) \rangle)$
(TT)	$T(\langle T(\langle p \rangle) \rightarrow p \rangle)$

Implication Principles

(MPC)	$(p_1 \wedge \dots \wedge p_n \rightarrow q) \rightarrow (T(\langle p_1 \rangle) \wedge \dots \wedge T(\langle p_n \rangle) \rightarrow T(\langle q \rangle))$
(SPC)	$(p \rightarrow q) \rightarrow (T(\langle p \rangle) \rightarrow T(\langle q \rangle))$
(Sub-In)	$p \leftrightarrow q \rightarrow T(\langle p \rangle) \leftrightarrow T(\langle q \rangle)$
(MPT)	$(T(\langle p_1 \rangle) \wedge \dots \wedge T(\langle p_n \rangle) \rightarrow T(\langle q \rangle)) \rightarrow (p_1 \wedge \dots \wedge p_n \rightarrow q)$
(SPT)	$(T(\langle p \rangle) \rightarrow T(\langle q \rangle)) \rightarrow (p \rightarrow q)$
(Sub-Out)	$T(\langle p \rangle) \leftrightarrow T(\langle q \rangle) \rightarrow (p \rightarrow q)$

It is reasonable to expect that our theory of the replacement concept(s) should include as many replacement principles—those like the above but formulated with the replacement concepts—as possible. With a single replacement for truth, one will end up with very few of these replacement principles. Consider the study by Harvey Friedman and Michael Sheard (1987) on just twelve principles that truth seems to obey. They document all the possible consistent subsets of these twelve principles. One lesson from their analysis is that trying to find a consistent subset of even the most basic principles we unreflectively take truth to obey is like navigating a minefield—there are so many hidden inconsistencies in seemingly innocuous combinations of just these twelve principles; figure 1 displays their results (any combination not labeled as

21. 'Imb' is short for 'imbibe'.

22. 'Exc' is short for 'excrete'.

Alethic Principles (from Friedman and Sheard)

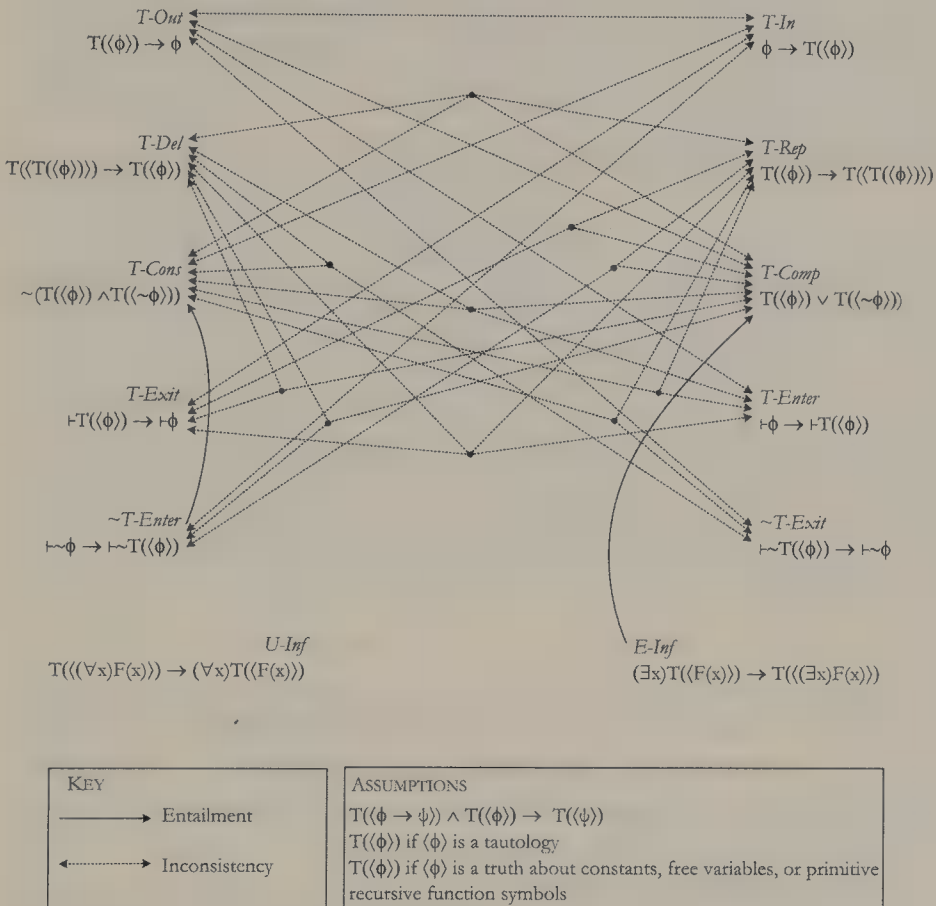


Figure 1.²³

inconsistent is consistent). It depicts seventeen distinct paradoxes just among these few principles.

And it only gets worse when one includes more truth-functional principles, quantification principles, and implication principles. Anyone who advocates replacing truth with a single concept would have to pick the best combination of replacement principles and give up anything like the rest of them. However, as I indicate below, when we replace truth with *two* concepts, we have the option of accepting some replacement principles that are formulated with one concept and some replacement

23. T-Intro, T-Elim, $\sim$ T-Intro, and $\sim$ T-Elim are to be read as derivability principles rather than inference rules; for example, T-Intro says that if p is derivable, then $T(\langle p \rangle)$ is derivable.

principles that are formulated with the other. In addition, this strategy allows for *hybrid* principles, which are formulated with both concepts. The theory of the replacement concepts I offer includes a replacement principle for every one of the principles listed above for truth. Such a thing is possible only when we replace truth with a team of concepts instead of a single concept.

Having argued that we should not replace truth with a single concept, we still need to decide on a team of replacement concepts. There are many options here, but one obvious way to go is: pick the smallest inconsistent collection of the most central of truth's principles and divide them up between replacement concepts. The most plausible candidate collection is: (T-In) and (T-Out).²⁴ So we should have one replacement concept that obeys an analogue of (T-In) but not the analogue of (T-Out) and another replacement concept that obeys an analogue of (T-Out) but not the analogue of (T-In). Inspired by Quine's comment that (T-In) encapsulates truth's function of semantic ascent, I call the concept that obeys (T-In) *ascending truth*. The other I call *descending truth*.

2.4. ADT

At this point, we have started characterizing our replacement concepts, but there is more work to be done. Descending truth obeys the principle

$$(D1) \ D(\langle p \rangle) \rightarrow p,$$

and ascending truth obeys the principle

$$(A1) \ p \rightarrow A(\langle p \rangle).$$

However, neither can obey the inverse principle on pain of contradiction. Nevertheless, the potential problems raised by the inverse principles arise only for a few sentences. It does no harm (and a lot of good) to let descending truth obey a restricted version of (T-In) and let ascending truth obey a restricted version of (T-Out). I use the term 'safe' for sentences for which ' $p \rightarrow D(\langle p \rangle)$ ' and ' $A(\langle p \rangle) \rightarrow p$ ' are correct.

Using the defining principles for safety and for ascending truth and descending truth, we can derive the following definition of safety:

$$(M2) \ S(\langle p \rangle) \leftrightarrow D(\langle p \rangle) \vee \sim A(\langle p \rangle)$$

24. I am excluding the substitution principle from the collection since it seems integral to being a predicate at all.

That is, a safe sentence is either descending true or not ascending true. Conversely, an unsafe sentence is both ascending true and not descending true. A consequence of this result is a clearer picture of the relation between descending truth and ascending truth. We know that if something is descending true, then it is ascending true, and if something is not ascending true, then it is not descending true. Further, some sentences are ascending true and not descending true. However, nothing is both descending true and not ascending true.

There is a tricky issue in specifying the relation between ascending truth and descending truth. Consider the relation between a sentence p , ' p is ascending true', and ' p is descending true'. p follows from ' p is descending true' but perhaps not vice versa; hence ' p is descending true' is a bit stronger than p . On the other hand, ' p is ascending true' follows from p but perhaps not vice versa; hence p is a bit stronger than ' p is ascending true'. Given these claims, 'it is not the case that p is descending true' is weaker than $\neg p$ and 'it is not the case that p is ascending true' is stronger than $\neg p$. So, what is the relation between ' p is ascending true' and 'it is not the case that $\neg p$ is descending true'? Further, what is the relation between ' p is descending true' and 'it is not the case that $\neg p$ is ascending true'? The most straightforward answer is that they are equivalent, which is the position I accept. That is, $A(\langle p \rangle) \leftrightarrow \neg D(\langle \neg p \rangle)$, and $D(\langle p \rangle) \leftrightarrow \neg A(\langle \neg p \rangle)$. Thus, ascending truth and descending truth are *dual* predicates.²⁵ Their relationship is the same as that that obtains between possibility and necessity, between permission and obligation, between consistency and provability, and so forth.

One of the most difficult problems facing our inchoate theory of ascending and descending truth is a theorem Richard Montague proved in 1963 that has had much more impact on the philosophical discussion of necessity than the discussion of truth. Montague (1963) proved that a theory of some predicate $H(x)$ with the following features is inconsistent:

- (i) All instances of ' $H(\langle p \rangle) \rightarrow p$ ' are theorems.
- (ii) All instances of ' $H(\langle H(\langle p \rangle) \rightarrow p \rangle)$ ' are theorems.
- (iii) All instances of ' $H(\langle p \rangle)$ ', where $\langle p \rangle$ is a logical axiom are theorems.
- (iv) All instances of ' $H(\langle p \rightarrow q \rangle) \rightarrow (H(\langle p \rangle) \rightarrow H(\langle q \rangle))$ ' are theorems.
- (v) Q (that is, Robinson arithmetic) is a subtheory.

25. Many thanks to Dana Scott who impressed upon me the importance of duality in the theory of ascending and descending truth.

Condition (v) is present to ensure that the language in which the theory is expressed has the ability to refer to its own sentences. The other four conditions are highly desirable for *descending* truth. On this reading, note that (i) is just the replacement for (T-Out), (ii) says that all instances of the replacement for (T-Out) are descending true, (iii) says that all tautologies are descending true, and (iv) says that descending truth is closed under modus ponens (that is, if a conditional is descending true, and its antecedent is descending true, then its consequent is descending true). Montague's theorem shows that if descending truth is a consistent concept, then it does not obey all four of these principles. Since I am taking the replacement for (T-Out) to be constitutive of descending truth, my options are to deny (ii), deny (iii), or deny (iv). Denying (ii) results in a theory of descending truth that entails that some of its axioms are not descending true, which is a version of a revenge paradox. I am committed to avoiding revenge paradoxes of any kind (more on this below in section 3.8.2). That leaves us with denying (iii) or denying (iv).

A recent result by Field (2006a) helps make this decision easier. He argues that the standard definition of validity (that is, necessary truth-preservation) is untenable in light of the alethic paradoxes because it is incompatible with every logical approach to the paradoxes.²⁶ In general, we have two prominent ways of thinking about validity: as the property of canons of good reasoning and as necessary truth-preservation. The lesson of Field's argument is that given any combination of a theory of truth and a logic, it is unacceptable that the canons of good reasoning necessarily preserve truth.²⁷ This argument of Field's is relatively new, and it is buried in a much more complex discussion of Gödel's second incompleteness theorem and formal theories of truth, so it has yet to generate much literature. However, I find it convincing, and this conclusion has an effect on my response to the problem posed by Montague's theorem because Field's considerations also sink any attempt to define validity in terms of descending-truth-preservation (assuming classical logic as a background). Thus, any theory of descending truth that accepts conditions (i), (ii), and (iii) of Montague's theorem has as a consequence that

26. Proponents of substructural views claim that they are not subject to Field's result; see Ripley (2012, forthcoming) and Beall and Murzi (forthcoming).

27. The effects of this split can be seen all over the literature on logical approaches to the paradoxes. For example, Maudlin (2004) defines validity in terms of truth preservation, and that leads him to claim that (T-In) and (T-Out) are valid on his theory. They are truth preserving according to his theory, but they are not canons of good reasoning according to his theory. See Field 2006b and Maudlin 2006 for discussion.

descending truth is not closed under some derivations. However, it is open to say that all logical truths are descending true. Thus, it makes the most sense to reject (iv) and accept (iii). As such, I stipulate that all classical first-order tautologies are descending true; it follows by Montague's theorem that descending truth is not closed under modus ponens.²⁸

Although I cannot justify or even describe all the other choices made to fill out the theory of ascending and descending truth, below is the theory, which I call ADT:²⁹

- (D1) $D(\langle p \rangle) \rightarrow p$
- (D2) $D(\langle \sim p \rangle) \rightarrow \sim D(\langle p \rangle)$
- (D3) $D(\langle p \wedge q \rangle) \rightarrow D(\langle p \rangle) \wedge D(\langle q \rangle)$
- (D4) $D(\langle p \rangle) \vee D(\langle q \rangle) \rightarrow D(\langle p \vee q \rangle)$
- (D5) $D(\langle p \rangle)$ if $\langle p \rangle$ is a logical truth (for definiteness, if $\langle p \rangle$ is a theorem of classical first-order predicate calculus).
- (D6) $D(\langle p \rangle)$ if $\langle p \rangle$ is a mathematical truth (for definiteness, if $\langle p \rangle$ is a theorem of Peano Arithmetic).
- (D7) $D(\langle p \rangle)$ if $\langle p \rangle$ is an axiom of ADT (that is, if $\langle p \rangle$ is an instance of D1–D6, A1–A6, M1–M4, or E1–E3).

- (A1) $p \rightarrow A(\langle p \rangle)$
- (A2) $\sim A(\langle p \rangle) \rightarrow A(\langle \sim p \rangle)$
- (A3) $A(\langle p \rangle) \vee A(\langle q \rangle) \rightarrow A(\langle p \vee q \rangle)$
- (A4) $A(\langle p \wedge q \rangle) \rightarrow A(\langle p \rangle) \wedge A(\langle q \rangle)$
- (A5) $\sim A(\langle p \rangle)$ if $\langle p \rangle$ is a contradiction (for definiteness, if $\langle \sim p \rangle$ is a theorem of classical first-order predicate calculus).
- (A6) $\sim A(\langle p \rangle)$ if $\langle p \rangle$ is a mathematical falsehood (for definiteness, if $\langle \sim p \rangle$ is a theorem of Peano Arithmetic).³⁰

- (M1) $D(\langle p \rangle) \leftrightarrow \sim A(\langle \sim p \rangle)$
- (M2) $S(\langle p \rangle) \leftrightarrow (D(\langle p \rangle) \vee \sim A(\langle p \rangle))$

28. A consequence of this decision is that neither ascending truth nor descending truth obeys the first assumption listed in figure 1. One might worry that this move undermines my argument against the proponent of a one-concept replacement (above). However, the results listed in figure 1 are merely intended to illustrate the vast number of inconsistencies among the principles that truth seems to obey. Even if closure under modus ponens is rejected, there are still many principles a one-concept replacement theorist would have to simply reject.

29. Formulating the principles of ADT with schematic variables is easier to read but not essential. There are some redundancies in this list. I discuss my reasons for this presentation below.

30. Axioms D5, D6, A5, and A6 hold for sentences that contain 'ascending true' and 'descending true'.

$$(M3) \ p \wedge S(\langle p \rangle) \rightarrow D(\langle p \rangle)$$

$$(M4) \ A(\langle p \rangle) \wedge S(\langle p \rangle) \rightarrow p$$

$$(E1) \ \text{If } s = t, \text{ and } \langle q \rangle \text{ results from replacing some occurrences of 's' with 't' in } \langle p \rangle, \text{ then } D(\langle p \rangle) \leftrightarrow D(\langle q \rangle).$$

$$(E2) \ \text{If } s = t, \text{ and } \langle q \rangle \text{ results from replacing some occurrences of 's' with 't' in } \langle p \rangle, \text{ then } A(\langle p \rangle) \leftrightarrow A(\langle q \rangle).$$

$$(E3) \ \text{If } s = t, \text{ and } \langle q \rangle \text{ results from replacing some occurrences of 's' with 't' in } \langle p \rangle, \text{ then } S(\langle p \rangle) \leftrightarrow S(\langle q \rangle).$$

One question that naturally arises is: is this theory consistent or are there new paradoxes hiding in here? Given Gödel's second incompleteness theorem, all we can hope for is a proof of relative consistency (that is, if some uncontroversial mathematical theory is consistent, then ADT is consistent), and because of the extreme difficulty with saying anything at all consistent about the alethic paradoxes, a relative consistency proof seems in order. Although relative consistency can be demonstrated via a proof of the soundness of ADT with respect to a particular semantics, it is too technical for this essay.³¹

I want to emphasize that ADT is a *rudimentary* axiomatic theory of ascending and descending truth, in the sense that any acceptable theory of ascending and descending truth has ADT as a subtheory. Other principles can be consistently added to it, but a discussion of their pros and cons will have to wait for another occasion. Certainly, one would want quantifier principles,³² and there might be other connective principles that are consistent with it.³³ Principles governing iterated attributions might be helpful and consistent with ADT as well.³⁴

One possibly surprising feature of ADT is that neither ascending truth nor descending truth need be preserved under logical deduction. That is, one can have a valid argument with all ascending true premises but a conclusion that is not ascending true; the same goes for descending truth. How disturbing is this result? Not very. By Field's result discussed above, no classical logical approach to the alethic paradoxes is consistent

31. See Scharp 2011.

32. For example, if a universal generalization is descending true, then all its instances are descending true.

33. It is obvious that we *cannot* have ' $A(\langle p \rangle) \wedge A(\langle q \rangle) \rightarrow A(\langle p \wedge q \rangle)$ ' (because the ascending liar and its negation are both ascending true) or ' $D(\langle p \vee q \rangle) \rightarrow D(\langle p \rangle) \vee D(\langle q \rangle)$ ' (because the disjunction of the descending liar and its negation is descending true).

34. For example, $D(\langle D(\langle p \rangle) \rangle) \leftrightarrow D(\langle \sim A(\langle \sim p \rangle) \rangle)$.

with the claim that valid arguments necessarily preserve truth. Thus, as part of an approach to the alethic paradoxes, one that advocates replacing truth with ascending and descending truth is no worse off than the other theories that are nontrivial in classical logic. Moreover, it is not the case that valid arguments might lead one seriously astray. At worst, if the premises of a valid argument are descending true, then its conclusion might not be descending true, but it will be ascending true.³⁵ This kind of thing will come up only in cases of unsafe sentences.

A consequence is that although all the *axioms* of ADT are descending true (by virtue of axiom schema D7), it is not the case that all *theorems* of ADT are descending true. In fact, it is easy to find theorems that are not descending true—I present some of these in the next subsection. It is because of this feature that there are some redundancies in the list of axiom schemata of ADT. By including all these axiom schemata, I ensure that all their instances are descending true. There is much more to be said about this topic, but space limitations prevent a more thorough discussion.

Earlier I advertised that this prescriptive theory includes a replacement principle for each of the principles it is commonly thought that truth obeys. First, note that in the list of principles for truth above, if p and q are safe, and one replaces each occurrence of ‘T’ with either ‘A’ or ‘D’, then the result is a theorem of ADT. That result is not all that exciting because it is restricted to safe sentences. However, for each of the above principles of truth, there is an unrestricted principle that is either a principle of ascending truth, a principle of descending truth, or a hybrid principle. One can simply look at the definition of ADT for examples of principles of descending truth or principles of ascending truth that are from that original list. The hybrid principles are less obvious; the following is a list of examples with the original principle of truth on the left and a hybrid principle that is a theorem of ADT on the right:

Principle of Truth

Hybrid Principle

($\wedge$ -Imb) $T(\langle p \rangle) \wedge T(\langle q \rangle) \rightarrow T(\langle p \wedge q \rangle)$	$D(\langle p \rangle) \wedge D(\langle q \rangle) \rightarrow A(\langle p \wedge q \rangle)$
($\vee$ -Exc) $T(\langle p \vee q \rangle) \rightarrow T(\langle p \rangle) \vee T(\langle q \rangle)$	$D(\langle p \vee q \rangle) \rightarrow A(\langle p \rangle) \vee A(\langle q \rangle)$
(SPC) $(p \rightarrow q) \rightarrow (T(\langle p \rangle) \rightarrow T(\langle q \rangle))$	$(p \rightarrow q) \rightarrow ((D(\langle p \rangle) \rightarrow A(\langle q \rangle))$
(SPT) $(T(\langle p \rangle) \rightarrow T(\langle q \rangle)) \rightarrow (p \rightarrow q)$	$(A(\langle p \rangle) \rightarrow D(\langle q \rangle)) \rightarrow (p \rightarrow q)$

35. Note that we cannot derive this result *in ADT* because of Gödel’s Second Incompleteness Theorem.

Nothing like a hybrid principle is possible unless one replaces truth with two or more concepts. If one follows the replacement strategy I have outlined here, then one can accept a replacement principle for each of the principles that truth seems to obey. I take that to be strong support for this prescriptive theory.

2.5. *Ascending and Descending Liars*

The theory of truth is the topic of the next section—here we deal only with *ascending truth* and *descending truth*. Thus, since liar sentences, Curry sentences, and Yablo sentences all contain truth predicates or falsity predicates, a discussion of them is reserved for section 3. Here I want to consider sentences like these that contain ‘ascending true’ or ‘descending true’. Consider the following sentences that are the analogues of liar sentences:

- (2) (2) is not ascending true.
- (3) (3) is not descending true.

Call (2) the *ascending liar* and (3) the *descending liar*. It is easy to show that (2) and (3) are each unsafe—that is, they are ascending true and not descending true. The standard argument in the liar reasoning uses both (T-In) and (T-Out). However since neither ascending truth nor descending truth obey both these rules, the standard argument is invalid.

Assume (3) is descending true.	Assume (2) is ascending true.
‘(3) is not descending true’ is descending true.	‘(2) is not ascending true’ is ascending true.
(3) is not descending true.	<i>(2) is not ascending true.</i>
Assume (3) is not descending true.	Assume (2) is not ascending true.
<i>‘(3) is not descending true’ is descending true.</i>	‘(2) is not ascending true’ is ascending true.
(3) is descending true.	(2) is ascending true.

The steps leading to the italicized sentences are invalid. In the argument on the left, the inference is from ‘(3) is not descending true’ to ‘“(3) is not descending true’ is descending true,” which is an instance of ‘if p, then D(⟨p⟩)’; this inference rule is not valid in general for descending truth. In the argument on the right, the inference is from ‘“(2) is not ascending true’ is ascending true’ to ‘(2) is not ascending true’, which is an instance of ‘if A(⟨p⟩), then p’; this inference rule is not valid in general for ascending truth. So neither of these sentences poses a problem for

ADT. Moreover, ADT implies that they are ascending true and not descending true (that is, unsafe).

Since Curry paradoxes and Yablo paradoxes follow the same pattern—they depend on applications of both (T-In) and (T-Out)—the results will be the same there. Those sentences are unsafe, and those arguments do not pose a problem for ADT.³⁶

2.6. *Empirical Unsafety*

It is common knowledge among those who work on the liar paradox that one can construct paradoxical sentences with the use of empirical predicates.³⁷ These sentences are paradoxical because of some empirical facts; if the facts had been different, they would not have been paradoxical. We can say that such sentences are *empirically paradoxical*.

Philosophers have known of empirical versions of the liar paradox since it first became an object of study over two millennia ago. For example, the predicate ‘is a complete sentence in section 2.6 of Scharp’s “Truth, the Liar, and Relativism” whose first letter is an “E”’ can be used to construct a version of the liar paradox.

Every complete sentence in section 2.6 of Scharp’s “Truth, the Liar, and Relativism” whose first letter is an ‘E’ is not true.

The fact that the previous sentence is the only complete sentence in this section of this essay to begin with an ‘E’ is an empirical fact about that sentence. If I had chosen to place it in a different section or if I had included some other sentences in this section, then it might not have uniquely satisfied that empirical predicate, and, thus it would not have predicated falsity of itself alone. Nevertheless, it seems obvious that this change would not have altered the sentence’s syntactic or semantic features.³⁸

A version of empirical paradoxicality survives in the switch to ascending and descending truth in the form of empirical unsafety. That is, whether a sentence is unsafe can depend on just about any

36. See Scharp 2011 for details.

37. Kripke (1975) emphasizes this point.

38. Did you look at every sentence in this section to confirm? If so, then you now have a great example of knowing the syntactic and semantic features of a sentence without knowing whether it is paradoxical. Think for a moment how bizarre it would be to say that while you were perusing the other sentences of this section looking for ‘E’s, you were learning about the syntax or the meaning of that one sentence.

empirical facts (consider replacing ‘true’ with ‘descending true’ or ‘ascending true’ in the displayed sentence above). Although this point might seem insignificant, it plays an important role in considerations about the correct theory of truth in the following section.

2.7. *Grasping Ascending Truth and Descending Truth*

One might worry that simply familiarizing oneself with the formal theory ADT is not enough to really grasp the concept of ascending truth and the concept of descending truth. It is, after all, just a collection of axioms. In this subsection, I aim to provide more grounding for the formal theory by doing three things: (i) sketching the way ADT should be physically interpreted, (ii) describing the ways ascending truth and descending truth serve truth’s expressive role, and (iii) showing how ascending truth and descending truth can play truth’s explanatory role in a theory of meaning.

2.7.1. Interpreting ADT

The guiding principle for interpreting ADT is that ascending truth and descending truth should be as close as possible to one another (each one thereby approximating the inconsistent concept of truth). Since the ascending truth values and descending truth values of sentences are different only for unsafe sentences, we can think of the guiding principle as saying that we should strive to minimize the set of unsafe sentences when interpreting ADT.

It is compatible with ADT that for a classical first-order language that contains no semantic vocabulary, ascending truth and descending truth coincide on all of its sentences. That is, none of its sentences are unsafe. That result holds regardless of what kinds of empirical claims or mathematical claims can be expressed in the language. So, in empirical or mathematical discourse, one can reason using either ascending truth or descending truth as if it were a truth predicate. If we add some way for the language to refer to its own sentences and add an ascending truth predicate, a descending truth predicate, and a safety predicate, then the language will contain some unsafe sentences. However, every sentence that is grounded (in something like Kripke’s [1975] sense) is safe.³⁹ That

39. I am using ‘grounded’ in the sense of *is such that ascending truth value and descending truth value are completely determined by the ascending truth values and descending truth values of sentences that have no occurrences of ‘ascending true’ or ‘descending true’*.

is, ascending truth attributions and descending truth attributions that eventually ground out in sentences of the original language are safe. In addition, many ungrounded sentences, for example, ‘every sentence is either ascending true or not ascending true’ and ‘no sentence is both descending true and not descending true’ are safe.⁴⁰ In short, only sentences that contain ‘ascending true’, ‘descending true’, or ‘safe’ might turn out to be unsafe, and even among those, only sentences that would be paradoxical if ‘true’ were substituted in for these terms might be unsafe.

These results add quite a bit to the understanding of ascending truth and descending truth that one acquires by simply perusing the axioms in ADT. For example, if *p* and *q* are grounded, then every principle on the above list of aletheic principles is valid when either ‘A’ or ‘D’ is uniformly substituted for ‘T’. Moreover, if *p* and *q* are grounded, then they have the same ascending and descending truth values. If one wants to delve deeper into the formal intricacies of ADT, then the next place to look is the semantics that are used to prove the relative consistency result.⁴¹ Unfortunately, presenting this material here would take us too far off track.

2.7.2. The Expressive Roles of Ascending and Descending Truth

It is essential to keep in mind that *I am not suggesting that we should stop using ‘true’*. For most purposes, it works fine, so there is no need to inform the public with flyers or public service announcements. However, in certain circumstances, for example, providing a semantics for an expressively rich language, one should use the replacement concepts. This situation is similar to what has happened with ‘mass’. People still use the concept of mass frequently—it is a useful inconsistent concept. However, there are situations in which it does not make sense to use ‘mass’, and one should instead use its replacements (for example, when calibrating the atomic clocks on GPS satellites). When should one stop using ‘mass’ and use ‘proper mass’ or ‘relativistic mass’ instead? When the difference between proper mass and relativistic mass is not negligible for the purposes at hand. We can say the same thing about truth—one should use ascending truth or descending truth instead of truth when the difference between them matters for whatever one is trying to accomplish. In practice, if one

40. This is not a trivial result, but I cannot pause to justify it here; see Scharp 2011.

41. See Scharp 2011, 2013, for details.

is not a semanticist, philosopher of language, or logician, then one will probably not ever need to use them.

When people find themselves in a position of wanting to use ‘ascending true’ or ‘descending true’ instead of ‘true’, they have a choice. Assume for simplicity that they are considering a single sentence, *p*. They can attribute ascending truth or descending truth to *p*. How should they decide? Remember, descending truth is stronger than ascending truth in the sense that if something is descending true, then it is ascending true, but not vice versa. If they attribute descending truth, and *p* turns out to be descending true, then their attribution is descending true, which is fine.⁴² If, however, *p* is unsafe (that is, ascending true but not descending true), their attribution is unsafe as well. That is not necessarily a bad thing—some unsafe sentences are still assertible, and their attribution is still ascending true.⁴³ If *p* is not ascending true, then their attribution is not ascending true, which should be treated as we treat an assertion we take to be false.⁴⁴ On the other hand, if they decide to attribute ascending truth, and *p* is descending true, then their attribution is descending true, which, again, is fine. Just as before, if *p* is unsafe, their attribution is unsafe, and if *p* is not ascending true, then their attribution is not ascending true. Thus, their decision about which predicate to use should be based on how strong a claim they want to make and whether they think that the sentence(s) in question might be unsafe. If they think that they might be unsafe, then they should use ‘ascending true’; if not, they should use ‘descending true’. Remember, this decision only arises in situations where the difference between ascending truth and descending truth is not negligible relative to the person’s interests—since the vast majority of sentences with which one is likely to be concerned are safe, this will rarely be an issue.

42. *p* is safe (because it is descending true by assumption), so ‘*p* is descending true’ is safe. If ‘*p* is descending true’ were not ascending true, then it would follow that *p* is not descending true. However, we already assumed that *p* is descending true. Thus, ‘*p* is descending true’ is descending true.

43. Explaining the relation between assertibility and ascending and descending truth is an important part of the project, but I do not have the space to devote to it here. See Scharp 2013.

44. *p* is safe. Assume for reductio that ‘*p* is descending true’ is ascending true. It follows that ‘*p* is descending true’ is descending true (because *p* is safe). Thus, *p* is descending true, which entails that *p* is ascending true. We assumed that *p* is not ascending true, so we have a contradiction. Thus, ‘*p* is descending true’ is not ascending true.

Consider how ascending truth and descending truth perform truth's expressive role. I mentioned earlier that truth predicates function as devices of endorsement and rejection. How well do ascending truth and descending truth perform these jobs? We know that p follows from ' p is descending true', so descending truth functions as a device of endorsement. If a person asserts 'the Flanders hypothesis is descending true', then he or she has thereby endorsed the Flanders hypothesis. On the other hand, ' $\sim p$ ' does not necessarily follow from ' p is not descending true'; thus, if one asserts 'the Flanders hypothesis is not descending true', one need not thereby have endorsed the negation of the Flanders hypothesis. Instead, for rejections, one would want to use ascending truth; if one asserts 'the Flanders hypothesis is not ascending true', then one has committed oneself to the negation of the Flanders hypothesis. Thus, descending truth serves as a device of endorsement, and ascending truth serves as a device of rejection.

2.7.3. Doing Semantics with Ascending and Descending Truth

So far, I have argued in detail that truth should be replaced for certain purposes not by one concept but by two—ascending truth and descending truth. Moreover, I presented the axiomatic theory ADT and explained how it handles liar-like sentences involving 'ascending true' and 'descending true'. I have also described how best to interpret ADT so that the set of safe sentences is maximized and discussed the way 'ascending true' and 'descending true' are to be used by speakers. One might want still more detail about how to make sense of these concepts.

In this subsection, I follow the broad outlines of Donald Davidson's strategy for giving a nonreductive theory of truth, except that I am interested in the theory of ascending truth and descending truth. Davidson (1990, 313–14) writes,

A theory of fundamental measurement of weight, for example, states in axiomatic form the properties of the relation between x and y that holds when x is at least as heavy as y ; this relation must, among other things, be transitive, reflexive, and nonsymmetric. A theory of preference may stipulate that the relation of weak preference has the same formal properties. But in neither case do the axioms define the central relation (x is at least as heavy as y , x is weakly preferred to y), nor instruct us how to determine when the relation holds. Before the theory can be tested or used, something must be said about the interpretation of the undefined concepts. The same applies to the concept of truth.

It is a mistake to look for a behavioristic definition, or indeed any other sort of explicit definition or outright reduction of the concept of truth. Truth is one of the clearest and most basic concepts we have, so it is fruitless to dream of eliminating it in favor of something simpler or more fundamental. Our procedure is rather this: we have asked what the formal properties of the concept are when it is applied to relatively well-understood structures, namely, languages.

Davidson's central methodological commitment is that in lieu of a reductive theory of truth, which he thinks is hopeless, he endorses an axiomatic theory (in Davidson's case, it is a Tarskian theory) and then aims to chart the connections between truth and other concepts, including meaning, belief, desire, and rationality. I accept this sort of methodology and the analogy Davidson draws between a theory of interpretation and a theory of fundamental measurement. I would love to present this sort of project for ascending and descending truth, but it would be impossible to do anything like this exercise in detail. Instead, let me say a bit about the connection between meaning and ascending truth and descending truth. Arguably, truth serves many other explanatory roles as well, but they will have to wait for some other occasion.⁴⁵ *Please note well:* this is not just an illuminating exercise—I use this framework in section 3 when I offer an assessment-sensitive semantic theory for our inconsistent concept of truth.

I am going to assume that the reader has some familiarity with formal semantics in the Montogovian and Kaplanian traditions. Because there is little agreement on the terminology, especially when it comes to more contentious views like nonindexical contextualism and assessment-sensitivity, this presentation follows the recent and influential treatment of formal semantics for natural language by Stefano Predelli (2005). Although any presentation of this topic is bound to generate controversy, I do not think that the details matter—I could have used any recent account.

Predelli is careful to distinguish between a *linguistic practice*, which consists of rational entities making noises and inscriptions in the course of their interactions with other rational entities, and an *interpretive system*,

45. Deflationists about truth usually think that truth should not play an explanatory role at all, but given the importance of truth conditional theories of meaning in linguistics (an empirical science), I find it hard to take this position seriously. Either way, deflationists about *truth* could accept that *ascending truth* and *descending truth* play explanatory roles in semantics in the way outlined below.

which is used as a tool by natural language semanticists to explain facts about the semantic properties of those noises and inscriptions. On Predelli's view, there is a layer of processing that occurs between the linguistic practice and the interpretive system. For this reason, interpretive systems do not take natural language sentences as input; instead, their inputs are complex structures that result from disambiguating sentences. I use Predelli's neutral term 'clause' for these items.⁴⁶ In addition, the interpretive system needs information about the context of utterance (for context-dependent expressions). Again, following Predelli, I use the term 'index' for the information that gets fed into the interpretive system and 'context' for the concrete environment in which the utterance is performed.⁴⁷ Just as interpretive systems accept only specific inputs, they produce special outputs. The goal is assigning truth conditions to sentences uttered in the linguistic practice, but there is an additional level of complexity between the output of the interpretive system and the assignment of truth conditions. Instead, the interpretive system outputs *t-distributions*, which are assignments of truth values to clause/index pairs (or propositions) relative to points of evaluation. The points of evaluation contain information like a possible world and a time.

In order to accommodate the assessment-sensitivity view of 'true' proposed in section 3, we need to distinguish between a presemantic theory, a semantic theory, and a postsemantic theory.⁴⁸ Interpretive systems are semantic theories—they take clause/index pairs as input and produce *t-distributions* as output. *Presemantic* theories take natural language utterances as input and produce clause/index pairs as output. Thus presemantic theories relate natural language utterances to semantic theory inputs. *Postsemantic* theories take *t-distributions* and indexes as input and produce truth values and truth conditions for natural language utterances. Hence, postsemantic theories relate semantic theory outputs to natural language utterances. In sum, we begin with a natural language utterance, run it through a presemantic theory to arrive at a clause/index pair, then use a semantic theory to compute a *t-distribution* for that

46. Some theorists deny that there is a distinction between clauses (that is, inputs to a formal semantic theory) and natural language expressions—nothing in my treatment hangs on it.

47. Note that contexts in this sense can be modeled using standard pragmatic theories; see Stalnaker 1970, 1974, 1978, 1998; Lewis 1983; and Roberts 2012 [1996], 2004, 2010.

48. I get the term 'presemantic' from Perry 2001 and 'postsemantic' from MacFarlane 2005a.

clause/index pair, and finally use a postsemantic theory on that t-distribution to generate truth conditions and a truth value for the natural language utterance with which we began. See figure 2 for a handy diagram.

Again, the terminology here is controversial. The distinction between semantic theory and postsemantic theory is *not* the distinction between a theory of linguistic content and a theory of pragmatic phenomena like implicature. Rather, the distinction between semantic theory and postsemantic theory is within the realm of linguistic content. It is required by any theory that distinguishes the proposition expressed by

Presemantic, Semantic, and Postsemantic Theories

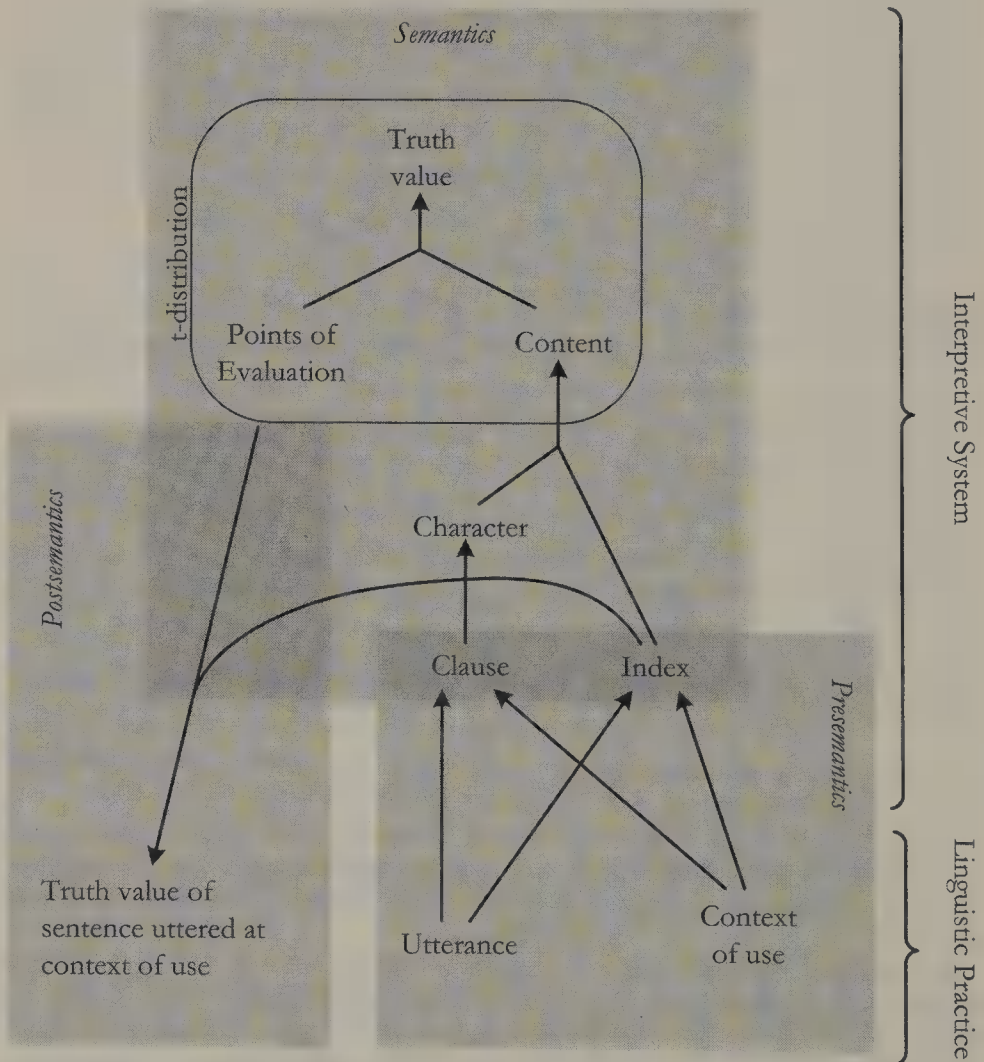


Figure 2.

a sentence (or the content of a linguistic expression) and the truth conditions of a sentence (or the extension of a linguistic expression). As I use the terms, contents, including propositions, are assigned by a semantic theory, and truth conditions and extensions are assigned by a postsemantic theory, which takes as input propositions and contents together with information from the context of utterance (for example, world) and possibly the context of assessment. Many people use ‘semantic theory’ to cover the totality of what I am calling presemantic theory, semantic theory, and postsemantic theory. However, nothing of substance turns on this terminological difference.

Let us now look at intensional semantic theories, which assign truth values to propositions. Following Kaplan (1989), we distinguish between two levels in the semantic theory. At the first level, expressions are assigned a character, while at the second level, the character/index pair is assigned a content. The output of the intensional semantic theory is an assignment of an extension to this content at every point of evaluation (in the case of sentences, their contents are propositions, and their extensions are truth values). As such we can think of characters as functions from indexes to contents, and we can think of contents as functions from points of evaluation to extensions. See figure 3 for a diagram.

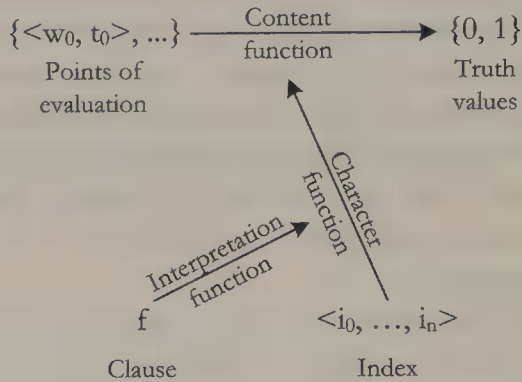
The output of an intensional semantics (a *t*-distribution) is an assignment of truth values to contents at points of evaluation. To get from this to a truth value and a truth condition for the utterance in question, we need a postsemantic theory. It is standard to use something like:

- (4) A sentence *p* is true in a context *u* iff the content assigned to the clause that represents *p* with respect to the index that represents *u* is true at the point of evaluation that represents the world and time of *u*.

Note that we have defined the truth of a sentence in a context in terms of the output of the semantic theory (more on this below).

After that brief introduction, we are ready to consider how to replace truth with ascending truth and descending truth in this framework. The concept of truth plays no role in the presemantics, so it is fine. The semantics assigns truth values to contents at points of evaluation, and the postsemantics provides a recursive definition of ‘sentence *x* is true in a context of use’ or ‘sentence *x* is true in a context of use from a context of assessment’ (depending on whether one admits assessment-sensitive expressions). I claim that truth is not really playing a role in the semantics, but it is in the postsemantics.

Intensional Semantics



Clause: a formula of some logical calculus

Index: n-tuple of information from the context of utterance

Point of evaluation: world/time pair

Interpretation: function from clauses to characters

Character: function from indices to contents

Content: function from points to truth values

Figure 3.

Begin with the postsemantics. Although the notions of truth at a context of use and truth at a context of use from a context of assessment are distinct from the concept of truth (that is, the latter is monadic, while the two former are polyadic), they are still susceptible to paradox. In particular consider the following sentences:

- (5) For all contexts of use u , (5) is not true in u .
- (6) For all contexts of use u and for all contexts of assessment a , (6) is not true in u from a .

These are both paradoxical (indeed, (5) is routinely used in revenge paradoxes for contextual approaches to the paradoxes).⁴⁹ This is good

49. See Juhl 1997.

evidence that these concepts are inconsistent (I am not going to argue this point further).

In order to avoid using a postsemantic theory that appeals to inconsistent concepts, we need a new postsemantic theory. There are two obvious ways to do this:

- (i) a postsemantic theory that takes as input the *t*-distribution from a semantic theory (whatever that turns out to be) and outputs the ascending truth value of a sentence in a context of use and the descending truth value of a sentence in a context of use.
- (ii) a postsemantic theory that takes as input the *t*-distribution from a semantic theory (whatever that turns out to be) and outputs the ascending truth value of a sentence in a context of use from a context of assessment and the descending truth value of a sentence in a context of use from a context of assessment.

If a language contains no assessment-sensitive terms at all, then (i) will be fine; otherwise (ii) is appropriate. Note that these theories make use of predicates that we have not seen yet: 'x is ascending true in context u', 'x is descending true in context u', 'x is ascending true in context u from context a', and 'x is descending true in context u from context a'. ADT, the theory of ascending and descending truth, is a theory of the one-place predicates 'x is ascending true' and 'x is descending true'; thus ADT does not serve as a theory of these new polyadic predicates. Although it might be helpful to develop a formal theory of them, I am not going to do so here. Note that one need not adopt a formal theory of 'x is true in context u' to be able to use it in semantics for natural language. For example, John MacFarlane (2005a) appeals to theories of assertion to help readers understand 'x is true in context u from context a', but he does not give even an intuitive theory of it, much less a formal theory. From what has been said so far about ascending truth and descending truth, these new predicates should not pose any real problems in understanding. At worst, one would have to say that they are implicitly defined by the postsemantic theory in question.

So much for replacing truth in the postsemantics; what about the semantic theory? A semantic theory outputs a *t*-distribution, which is an assignment of a truth value to a proposition at each point of evaluation. Does this use of truth need to be replaced as well? I do not think so. If we look at the way the semantic theory works, we can see that it is powered by truth-in-a-model, not truth. Truth-in-a-model is a mathematical concept, and it is not the same as truth. For discussion, fix a model $\mathfrak{M}$. The claim

that some clause is true-in- $\mathcal{M}$ is fully representable in set theory as a mathematical function from one set-theoretic entity to another. As long as the relevant mathematical theory (for example, ZFC) is consistent (and we have no reason to think it is not), there is no problem whatsoever with truth-in- $\mathcal{M}$. It is not an inconsistent concept.

Still, one might ask, what about paradoxes with truth-in- $\mathcal{M}$? As long as our language has the relevant mathematical locutions, we can formulate a sentence such as:

(7) (7) is not true-in- $\mathcal{M}$.⁵⁰

However, there is no reason to think that (7) is paradoxical. For one, (T-In) and (T-Out) fail for true-in- $\mathcal{M}$. That is, there is no reason to think that ‘p is true-in- $\mathcal{M}$ ’ follows from p since there is no reason to think that p is even representable in the language for which $\mathcal{M}$ is a model, and even if it is representable, there is no reason to think that p would be true-in- $\mathcal{M}$. For example, $\mathcal{M}$ might assign some bizarre meaning to p. In addition, ‘p is true-in- $\mathcal{M}$ ’ does not entail p since there is no reason to think that $\mathcal{M}$ represents the way the world is. Obviously, representing the world is the notion that would be needed to get (T-Out), but the semantics does not appeal to it. Consider another sentence:

(8) For all $\mathcal{M}$, (8) is not true-in- $\mathcal{M}$.

Does (8) pose a problem? No. Quantifying over set-theoretic entities is a complicated business, and explaining why (8) is benign would take us too far afield into the technical details of mathematical logic. It should be sufficient to note that (8) is a mathematical claim, pure and simple. As long as set theory (ZFC will do) is consistent, we know that (8) does not pose a problem.

Still, a question remains: how do we get from a t-distribution (as the output of the semantic theory) to an assignment of ascending truth conditions and descending truth conditions? There are a couple of ways to accomplish this: (i) add an extra slot to the points of evaluation that selects whether we are evaluating the proposition in question for ascending truth or descending truth, or (ii) define ascending-truth-in-a-model and descending-truth-in-a-model and assign values for each indepen-

50. Actually, this is harder than it seems. It is difficult to construct a model for a language that makes one of the singular terms in that language refer to that model. The problem comes in letting the model be a member of the domain that is an element in that very model.

dently to each proposition (for example, assign either 0 or 1 for the ascending-truth-in-a-model value and either 2 or 3 for the descending-truth-in-a-model value). Although (ii) is more elegant and probably the correct way to proceed, it would require a lengthy detour through the semantics for ADT, which is too complex and technical for this essay; it would also treat a proposition as a *relation* between the set of points of evaluation and $\{0, 1, 2, 3\}$ instead of as a *function* from the set of points of evaluation and $\{0, 1\}$, which would need some discussion. So, in this treatment, I follow the vulgar but accessible option (i) and leave (ii) for future work.

The extra parameter to be added to the points of evaluation has only two possible values: A or D. At the point of evaluation $\langle w_0, t_0, A \rangle$, where w_0 and t_0 are some world and some time, the proposition in question is assigned 0 if it is not ascending true at w_0 and t_0 , and it is assigned 1 if it is ascending true at w_0 and t_0 . At the point of evaluation $\langle w_0, t_0, D \rangle$, the proposition in question is assigned 0 if it is not descending true at w_0 and t_0 , and it is assigned 1 if it is descending true at w_0 and t_0 . I call this parameter the *aletheic value parameter*. The aletheic value parameter is used by the postsemantic theory. To arrive at the ascending truth value for the sentence uttered, one uses the point of evaluation with the relevant world, relevant time, and the ascending aletheic value. To arrive at the descending truth value for the sentence uttered, one uses the point of evaluation with the relevant world, relevant time, and the descending aletheic value. If this presentation is too abstract, the example given in the next section should help illustrate how this proposal works.

In sum, I have shown how to replace truth with ascending truth and descending truth in contemporary semantics. I have suggested that truth should be replaced in the postsemantic theory with ascending truth and descending truth; that is, the postsemantic theory outputs the ascending truth value of the sentence uttered in the context of use and the descending truth value of the sentence uttered in the context of use. Accordingly, the points of evaluation have an extra parameter—the aletheic value parameter—which controls whether the proposition in question is being evaluated for ascending truth or descending truth.

Before moving on, let us consider an objection to this whole project of replacing truth in the combination of a presemantic, semantic, and postsemantic theory (a *PSP theory* from here on) with ascending truth and descending truth. It is common to motivate truth-conditional PSP theories by saying that those of us who are competent with a natural language have certain implicit or explicit knowledge. We know the syntactic rules

that determine which strings count as well-formed sentences of that language, and we know the semantic features (for example, referents and extensions) of its basic linguistic expressions. We use this knowledge to decide which sentence of the language to utter in a given situation and how to interpret its sentences uttered by others. A PSP theory for this language explains our linguistic competence because it is our knowledge of this very theory that allows us to count as competent with the language in question. This sort of motivation presupposes that we possess and use the very concept of truth at the heart of a truth-conditional PSP theory for our language.⁵¹ Ordinary speakers clearly do not possess the concepts of ascending truth and descending truth, so if this motivation for PSP theories is correct, then the enterprise of replacing truth in such a theory with ascending truth and descending truth is a nonstarter. No PSP theory that utilizes ascending truth and descending truth instead of truth could be what most competent speakers of a language know.

Given the emphasis on a Davidsonian approach to the nature of truth described at the beginning of this subsection, it should not come as a surprise that I reject the particular motivation for truth-conditional PSP theories on which the objection is based. Instead of thinking of PSP theories as describing a person's psychological states that are responsible for his or her semantic competence, I prefer a Davidsonian view of their empirical significance. That is, if a person knew a PSP theory, then he or she would be able to understand the language in question. The evidence that supports such a theory comes from the person's behavior (verbal and otherwise) toward objects in the environment in accord with the procedure used by Davidson's hypothetical radical interpreter (see Davidson 1973). The theory itself describes certain aspects of what rational entities must be like for them to be able to interpret one another. I prefer to emphasize the aspects of Davidson's views that are most akin to the theory of fundamental measurement (mentioned at the beginning of this subsection). Probably the most familiar theory of this sort is presented and defended in Robert Matthews (2007), *The Measure of Mind*.⁵² Of course, nothing hangs on adopting this kind of justification of truth-conditional PSP theory in particular. There has been a vigorous debate about the relation between PSP theories and semantic competence; the Davidsonian

51. See Higginbotham (1991), Larson and Segal (1995), Antony (1997), and Barber (2001) for defenders of this view.

52. See Scharp 2013, chap. 7, for more details on this interpretation of formal semantics.

view and the cognitivist view expressed in the objection are but two of the players in this game. Some hold that PSP theories systematize and idealize what speakers actually know, while others claim that PSP theories describe objective semantic facts that are independent of speaker psychology.⁵³ Any of these views other than the cognitivist one expressed in the objection is compatible with the replacement strategy offered here. I happen to think that there are good reasons to reject the cognitivist view, but exploring this issue further would take us too far afield.

3. Truth ■ Assessment Sensitive

In the previous section, I introduced the idea that truth is an inconsistent concept and recommended that it be replaced with ascending truth and descending truth for certain purposes. I presented ADT, the theory of ascending and descending truth, as a rudimentary theory of these two replacement concepts (along with the concept of safety). ADT constitutes a prescriptive theory in the sense that it proposes a change to our conceptual scheme and natural language. Again, I am not suggesting that we stop using our truth predicate—only that we stop using it when our interest in using it is outweighed by the problems it causes (that is, the alethic paradoxes). In these cases, we should use the replacement concepts.

This prescriptive theory is only one part of an adequate inconsistency approach to the alethic paradoxes. The other major component should be a descriptive theory of our inconsistent concept of truth. One of the central theses of the approach I advocate is that the descriptive theory should not appeal to the concept of truth—otherwise we would be using an inconsistent concept to explain that very inconsistent concept, which is a recipe for disaster.⁵⁴ In this section, I offer a descriptive theory of truth that treats truth predicates as assessment sensitive, which is an endorsement of semantic relativism. The descriptive theory uses the semantics described at the end of the last section that appeals to ascending truth and descending truth.

3.1. *MacFarlane on a Semantics for Confusion*

Confusion occurs when a person thinks there is one thing (or one kind of thing), but there are really two (or more). To return to an earlier

53. See Dummett (1991) for the former and Devitt (1996, 2006) for the latter.

54. See Scharp 2007 for an argument.

example, after the late 1600s, but prior to the advent of relativistic mechanics, we thought that there is one physical quantity, mass, had by physical objects. However, in the early twentieth century, we realized that there is no physical quantity that obeys the principles we took mass to obey; instead, there are two physical quantities that are somewhat similar to mass: relativistic mass and proper mass. A person who lived in the 1800s and accepted Newtonian mechanics was confused—he or she thought that there is one quantity, mass, but instead there are two. We can say that the concept of mass is a confused concept, where the concept of mass is the concept expressed by ‘mass’. It is confused in the sense that it purports to pick out a single quantity, when in reality, there is no such thing—instead there are two quantities that are relevantly similar to what that single quantity would be like if it existed. Throughout the rest of this work, I use ‘confusion’ in this technical sense, which is considerably more specific than the ordinary sense of the word.

Once one accepts that truth is an inconsistent concept *and* that it should be replaced by ascending truth and descending truth, it is natural to think that the concept of truth is confused. It turns out that there is no property of truth, at least, if by ‘truth’ we mean a property that satisfies all of what I have claimed to be the constitutive principles of the concept of truth. Instead, there are two properties, ascending truth and descending truth.

I mention confusion here because John MacFarlane has suggested that words expressing confused concepts are assessment sensitive. MacFarlane quotes the following passage from Joseph Camp (2002, 125): “When one first thinks about ontological confusion, it is natural and intuitively plausible to talk in terms of perspectival truth. One wants to say: ‘what the confused person thinks may be true from one perspective but false from another perspective; or it may be true from both perspectives, or false from both’. Perspectival truth must replace truth simpliciter when one evaluates a confused belief.” MacFarlane (2007b, 708) then writes, “In recent work, I have suggested giving significance to perspectival truth by embedding it in a larger theory of language, specifically in a normative account of what it is to make an assertion. I would like to propose, very tentatively, that this kind of framework might be a better home for a ‘semantics of confusion’ than the multivalued, epistemic semantics Camp advocates.” MacFarlane does not give any details, and to my knowledge no one else has advocated anything like this. In the remainder of this essay, I work out the details of MacFarlane’s proposal as it applies to the particular case of ‘true’, using the new semantic

framework described above, where truth has been replaced by ascending truth and descending truth.⁵⁵

3.2. An Assessment-Sensitivity Theory of Truth

There are several things one might mean by ‘semantic relativism about truth’. Some use the term ‘semantic relativism’ for two separate views, nonindexical contextualism and assessment-sensitivity.⁵⁶ The latter is more radical than the former. Both require an extra parameter in each point of evaluation, but assessment-sensitivity views also require a new postsemantic theory that outputs the truth value of the sentence in a context of utterance from a context of assessment, instead of outputting the truth value of the sentence in a context of utterance. That is, postsemantics for an assessment-sensitivity view requires a three-place predicate ‘x is true in u from a’, whereas the postsemantics for nonindexical contextualist views needs only the more familiar ‘x is true in u’. In this subsection, I present my preferred semantics for the truth predicate, which is an assessment-sensitivity view. Later, I contrast it with several other options, including a nonindexical contextualist view.

Before we begin, it is worth a reminder that the semantic theory and postsemantic theory proposed do not use truth at all. Instead, they use ascending truth and descending truth, in accordance with the policy of *not* using the inconsistent concept of truth in a descriptive theory of truth (the semantic theory also uses the mathematical concept of truth-in-a-model).

Consider a linguistic practice in which rational agents utter sentences that contain truth predicates; for ease of exposition, we can take them to speak English. For each one of their utterances, the presemantic theory assigns a clause/index pair. The clause represents the sentence and carries information about the sentence’s syntactic structure or logical form—it treats ‘x is true’ as a univocal one-place predicate; the index represents the context of the utterance.

55. See also Pinillos 2010 for a semantic relativist treatment of ‘duration’, which in a Newtonian framework is arguably an inconsistent concept since it seems to presuppose absolute time.

56. Recanati (2007) and Richard (2008) endorse nonindexical contextualism for certain discourses and think of it as a species of semantic relativism; MacFarlane (2005a, 2009, forthcoming), on the other hand, thinks that nonindexical contextualism is not a kind of relativism—only assessment-sensitivity fits that description.

The semantic theory assigns a character to ‘true’, but this character ignores the index since ‘true’ is not an indexical.⁵⁷ That is, the content assigned to ‘true’ is invariant across indexes. From the index and the character of the expressions in the clause, the semantic theory assigns a proposition to the clause. The content of ‘true’, of course, contributes to the proposition assigned to the entire sentence. The last step for the semantic theory is to generate a *t*-distribution for the proposition. Recall that a *t*-distribution is an assignment of a truth value to the proposition for each point of evaluation; however, the term ‘truth value’ might promote misunderstanding here since the semantic theory uses truth-in-a-model. So, from here on, I use ‘*tM*-value’. Note too that because our postsemantic theory outputs ascending truth value and descending truth value (instead of truth value), our semantic theory already needs an alethic value parameter in its points of evaluation. That feature stems from replacing truth with ascending truth and descending truth—it has nothing to do with any kind of semantic relativism or specific view on the semantics for the truth predicate.

The specifically relativist feature of the semantics is that the points of evaluation have an additional parameter, which I call the *alethic standard*. Thus, the points of evaluation are $\langle \text{world, time, alethic value, alethic standard} \rangle$ quadruples. The alethic standard provides a “reading” of the occurrences of the truth predicate in the sentence in question. There are two possibilities for the alethic standard parameter: the ascending standard (S_A) and the descending standard (S_D). The ascending standard interprets the truth predicate as an ascending truth predicate, while the descending standard treats the truth predicate as a descending truth predicate. For example, a sentence ‘*p* is true’ is assigned *tM*-value 1 at points of evaluation $\langle \text{world, time, alethic value, } S_A \rangle$ iff ‘*p* is ascending true’ would be assigned *tM*-value 1 at that world, time, and alethic value; it is assigned *tM*-value 1 at points of evaluation $\langle \text{world, time, alethic value, } S_D \rangle$ iff ‘*p* is descending true’ would be assigned *tM*-value 1 at that world, time, and alethic value.

It is crucial to keep the *alethic value* parameter and the *alethic standard* parameter distinct—the former determines whether the post-semantics is assigning an ascending truth value or a descending truth value to the sentence in question, while the latter determines for sentences containing ‘true’ whether such occurrences are read by the

57. Throughout, I use ‘indexical’ in the general sense that includes any expression whose content depends on the context of utterance.

semantic theory as an ascending truth predicate or a descending truth predicate. If we return to the example just given and take w_0 to be a world and t_0 to be a time, then the semantics behaves in the following way:

- (i) 'p is true' gets $t\mathcal{M}$ -value 1 at $\langle w_0, t_0, A, S_A \rangle$ iff 'p is ascending true' is ascending true at w_0 and t_0 .
- (ii) 'p is true' gets $t\mathcal{M}$ -value 1 at $\langle w_0, t_0, A, S_D \rangle$ iff 'p is descending true' is ascending true at w_0 and t_0 .
- (iii) 'p is true' gets $t\mathcal{M}$ -value 1 at $\langle w_0, t_0, D, S_A \rangle$ iff 'p is ascending true' is descending true at w_0 and t_0 .
- (iv) 'p is true' gets $t\mathcal{M}$ -value 1 at $\langle w_0, t_0, D, S_D \rangle$ iff 'p is descending true' is descending true at w_0 and t_0 .

The *third* slot is the *aletheic value* parameter; it has nothing to do with semantic relativism and is present because we have replaced the concept of truth with the concepts of ascending truth and descending truth in the semantics and postsemantics: 'A' means that the semantics is specifying the ascending truth value of the proposition in question, and 'D' means that the semantics is specifying the descending truth value of the proposition in question. The *fourth* slot is the *aletheic standard* parameter; it is present because 'true' is assessment sensitive: 'S_A' means that the semantics treats 'true' in the sentence 'p is true' as if it were 'ascending true', and 'S_D' means that the semantics treats 'true' in the sentence 'p is true' as if it were 'descending true'. The biconditionals (i) – (iv) display the part of the t-distribution for 'p is true' at the particular world and time in question. In what follows, I use 'v' as an aletheic value variable and 's' as an aletheic standard variable.

An assessment-sensitive theory of truth also needs a postsemantics that takes t-distributions as input and outputs both ascending truth value in a context of use from a context of assessment and descending truth value in a context of use from a context of assessment. Here is a suggestion for that kind of postsemantic theory (letting u be a context of use, a be a context of assessment, i_u be the index representing u , and i_a be the index representing a):

- (9) A sentence p containing 'true' is *ascending true* in u from a iff the content assigned to the clause representing p with respect to i_u gets $t\mathcal{M}$ -value 1 at the point of evaluation $\langle w, t, v, s \rangle$, where w and t are the world and time of i_u , v is the *ascending* aletheic value parameter, and s is the aletheic standard from i_a .
- (10) A sentence p containing 'true' is *descending true* in u from a iff the content assigned to the clause representing p with respect to i_u gets

$\mathfrak{M}$ -value 1 at the point of evaluation $\langle w, t, v, s \rangle$, where w and t are the world and time of i_u , v is the *descending* aletheic value parameter, and s is the aletheic standard from i_a .

Notice that even though there are aletheic value slots and aletheic standards slots in the points of evaluation, they do not contribute to the content of the sentence in the context of utterance. Instead, the aletheic value slot determines which kind of truth conditions are being given (ascending or descending) for the sentence in question, and the aletheic standard slot determines the reading of the truth predicate in p and is picked up from the context of assessment. In short, the postsemantics has two inputs—the t -distribution from the semantics and the index from the presemantics—and two outputs—the ascending truth value of the sentence at the context of use from the context of assessment and the descending truth value of the sentence at the context of use from the context of assessment. The postsemantics focuses on two points of evaluation from the t -distribution: (i) the world of the context of use, the time of the context of use, the *ascending* aletheic value, and the aletheic standard from the context of assessment, and (ii) the world of the context of use, the time of the context of use, the *descending* aletheic value, and the aletheic standard from the context of assessment. The former determines the *ascending* truth value of the sentence at the context of use from the context of assessment, and the latter determines the *descending* truth value of the sentence at the context of use from the context of assessment.

Arguments whose sentences contain ‘true’ are assessed for validity in a natural way—an argument whose premises are the members of the set G and whose conclusion is p is valid iff for every point of evaluation e , if all members of G are assigned $\mathfrak{M}$ -value 1 at e , then p is assigned $\mathfrak{M}$ -value 1 at e . Because all the points of evaluation are classical, the resulting logic of the truth predicate is classical as well.

To sum up: the semantics and postsemantics for ‘true’ given here differ from those for a familiar, univocal, invariant predicate in two ways. First, rather than giving us the truth values for sentences containing ‘true’, the theory gives us the ascending truth values and the descending truth values for sentences containing ‘true’. It does this because we do not want the theory to use the inconsistent concept of truth, so it uses the replacement concepts instead. This feature has nothing to do with semantic relativism, and it is implemented by the aletheic value parameter in the points of evaluation (it could have been implemented differently). Second, rather than giving us the ascending truth value of

a sentence containing ‘true’ in a context of use and the descending truth value of a sentence containing ‘true’ in a context of use, the theory gives us the ascending truth value of a sentence containing ‘true’ in a context of use *from a context of assessment* and the descending truth value of a sentence containing ‘true’ in a context of use *from a context of assessment*. It does this because it treats ‘true’ as assessment sensitive. This feature is the semantic relativist aspect of the theory, and it is implemented by the alethic standard parameter in the points of evaluation.

3.3. Other Options

This subsection is devoted to comparing and contrasting the descriptive theory that takes ‘true’ to be assessment sensitive from the previous subsection with some other options and to giving reasons in favor of the assessment-sensitivity view. I consider the following five options: (i) ‘true’ is univocal, invariant, and determinate, (ii) ‘true’ is referentially indeterminate, (iii) ‘true’ is ambiguous, (iv) ‘true’ is an indexical, and (v) ‘true’ is a nonindexical context-dependent expression.

Option (i): truth is univocal, invariant, and has a determinate extension. Given the central claim of an inconsistency approach to the alethic paradoxes—that truth is an inconsistent concept—it is rather difficult to see how ‘true’, the predicate of English that expresses it, could be univocal, invariant, and have a determinate extension. What would its extension be? Any answer to this question is bound to be problematic since the reasoning for, say, the liar paradox can be cast as an argument that a liar sentence is in the extension of ‘true’ iff it is not. One could sidestep this problem by arguing that the truth predicate obeys (T-In) but not (T-Out) and so is like what has come to be known as a classical glutty truth predicate, or one could say that it obeys (T-Out) but not (T-In) and so is like what is called a classical gappy truth predicate.⁵⁸ However, it is difficult to see how one would argue for either one of these views over the other. If both (T-In) and (T-Out) are constitutive of truth, which is a central claim of the inconsistency approach, why would one of them hold and the other not? I do not see much promise for answering this

58. These terms are from Field 2008; see section 3.8.4 for details. Alternatively, one could argue that either the property of being ascending true or the property of being descending true (but not both) is a reference magnet; see Lewis 1983 [1979], Williams 2007, Hawthorne 2007, Sider 2009, and Chalmers 2012 for discussion of reference magnets. Although I reject the reference magnet view, arguing against it here would take us too far off topic.

question or for an inconsistency view that takes truth to be univocal, invariant, and have a determinate extension.⁵⁹ Instead, it seems to me that a predicate that expresses an inconsistent concept is bound to display some kind of semantic pathology or have some surprising semantic features. In addition, it seems that the replacement concepts, ascending truth and descending truth, should play some role in our understanding of the truth predicate.

Option (ii): ‘true’ is univocal and invariant, but it has a divided (indeterminate) reference. Before Hartry Field’s conversion to deflationism, he defended a particular kind of correspondence theory of truth (Field 1972) and introduced a phenomenon he dubbed “referential indeterminacy” (Field 1973). In fact, the mass example that has guided my discussion of inconsistent concepts is adapted from Field’s treatment. He suggests that ‘mass’ partially denotes relativistic mass and partially denotes proper mass. The semantics he suggests for ‘mass’ relies on two distinct interpretations of the language to which ‘mass’ belongs. One interpretation assigns ‘mass’ to the quantity of relativistic mass, and the other assigns it to the quantity of proper mass. It is straightforward to assign truth values to sentences containing ‘mass’ relative to each interpretation. That is, if we replace ‘mass’ in a sentence with ‘relativistic mass’ and the resulting sentence is true, then the original is true relative to the first interpretation; likewise for ‘proper mass’ and the second interpretation. Field (1973, 477–78) claims that a sentence containing ‘mass’ is determinately true if it is true on both interpretations, determinately false if it is false on both interpretations, and neither determinately true nor determinately false if the interpretations differ.

Field does not say how to define validity for arguments whose sentences contain an expression that is referentially indeterminate, but his method for assigning determinate truth and determinate falsity has come to be known as *supervaluation*, and the most familiar supervaluation consequence relation takes it that an argument whose premises constitute

59. An error theory of truth would also fall into this category—it implies that all atomic sentences with an occurrence of ‘true’ are false. Aside from being exceedingly implausible, the implications of this theory for the validity of arguments whose sentences contain ‘true’ are hard to live with. For example, it implies that any argument whose conclusion is an atomic sentence with ‘true’ is unsound. Moreover, it implies that any argument consisting entirely of atomic sentences with ‘true’ is valid (it is impossible that its premises are true and its conclusion is false). We can do much better than that. See Boghossian 2006 for discussion of error theories for (what I am calling) inconsistent concepts.

a set G and whose conclusion is a sentence p is valid iff if all the members of G are true in each interpretation, then p is true on each interpretation. Unfortunately, it is well known that if the language in question has a predicate expressing determinate truth (that is, truth in both interpretations in our case), then the logic of this language is not entirely classical. In particular, reasoning by cases, conditional proof, and reductio ad absurdum fail.⁶⁰ I stated earlier that I take it as a methodological assumption that an inconsistency approach should be compatible with classical logic. Thus, when paired with this version of supervenience, the indeterminacy option is unacceptable.

Instead, one might define validity differently (the above definition is often called *global*, and the following is called *local*): the argument with premises G and conclusion p is valid iff for each interpretation I , if all the members of G are true in I , then p is true in I . This version of supervenience has the benefit of being compatible with classical logic. It is also strikingly similar to the assessment-sensitivity theory that is the main focus of this essay. Notice that the liar reasoning is diagnosed as breaking down at the same point on both theories—if (1) is a liar sentence (that is, ‘(1) is not true’), then the instance of (T-In), ‘if (1) is not true, then (1) is true’ is not valid on either the local supervenience view or the assessment-sensitivity view. Likewise, the instance of (T-Out), ‘if (1) is true, then (1) is not true’, is also not valid on either account. Moreover, on each view, the truth predicate has no determinate extension. However, the local supervenience view by itself says nothing about the content of the truth predicate in the sense of content given by an intensional semantic theory. Indeed, the local supervenience view, conceived of as a logic for the truth predicate, is identical to the logic for the truth predicate specified by the assessment-sensitivity view. I am not going to prove this formally since it would require stating both views in considerable formal precision, which would not be worth the effort. Instead, once one recognizes that the alethic standards in the points of evaluation of the assessment-sensitivity view function in exactly the same way as the interpretations in the local supervenience view and that the definitions of validity are identical, it should be obvious that the theories are essentially the same. What the assessment-sensitivity view offers in addition to the logic of the truth predicate is an intensional semantics for the truth predicate. The

60. This is a well-known, although not uncontested, result. See Asher, Dever, and Pappas 2009 for details, and see Williams 2008 for an alternative view. In Field 2008, this is called a weakly classical logic.

interpretations used by the local supervaluation view are given empirical significance by the assessment-sensitivity theory as alethic standards that function in contexts of assessment. Thus, the local supervaluation view and the assessment-sensitivity view are not competitors. One can think of local supervaluationism as the logical core of the assessment-sensitivity theory of truth.⁶¹

Option (iii): ‘true’ is ambiguous. ‘True’ sometimes means *ascending true* and sometimes means *descending true*. There are many ambiguous expressions in English (‘bank’ is a common example), and they are usually treated as having two or more distinct meanings. When an ambiguous word is used in a conversation, the participants must disambiguate it—determine which meaning the speaker intended it to have. The intended meaning is usually obvious, and we disambiguate words frequently without even noticing it by using contextual cues. In more technical terms, the semantics has two predicates that we might label ‘true₁’ and ‘true₂’, which are treated as totally distinct. They are assigned different characters, they have different contents, and they make different contributions to the propositions expressed by sentences in which they occur. The presemantic theory has to disambiguate ‘true’ based on the context of utterance so as to select either ‘true₁’ or ‘true₂’ as the clause that represents it. One of these is assigned the same character as ‘ascending true’ and the other is assigned the same character as ‘descending true’.

There are several serious problems with the ambiguity option. First, linguists have compiled several tests for ambiguity, and ‘true’ fails all of them.⁶² For example, I cannot imagine a situation in which it would be felicitous to utter ‘p is true, but p is not true’, which is evidence against ‘true’ being ambiguous.⁶³ I am not going to go through all these tests—interested readers can investigate for themselves—mostly because there are, to my mind, more serious problems with this kind of theory. Second, many users of ‘true’ have no idea that it expresses an inconsistent concept, and the vast majority of them have never heard of ascending truth or descending truth, so it is pretty implausible to suggest that a speaker using ‘true’ has to intend that it means either *ascending true* or *descending true*.

61. There is much more that could be said about the relation between the two theories, but space does not permit a detailed discussion. See Kremer and Kremer 2003; Varzi 2007; Williams 2008; Cobreros 2011a, 2011b; and Fara 2011 for more on supervaluationism.

62. See Zwicky and Saddock 1975, Cruse 1986, Atlas 1989, Gillon 2004, and Kennedy 2010 on ambiguity and ambiguity tests.

63. Obviously this test presupposes that dialetheism is unacceptable.

Remember, the theory we select for ‘true’ should work for people who are totally unfamiliar with the liar paradox and any other reason for thinking that there is a problem with truth. A defender of the ambiguity view might suggest that something other than the speaker’s intuitions determines which meaning gets attached to ‘true’. I think that suggestion is implausible, but it incurs a further problem that I present in the next subsection; roughly, the context of utterance need not have enough information to disambiguate ‘true’.

Option (iv): contextualism about ‘true’. The idea here is that ‘true’ is univocal, the presemantics assigns a single clause to represent ‘true’, and the semantics assigns a single character to this clause. However, the character takes input from the index that represents the context of the conversation in question. The content of ‘true’ as used in a particular context is either *ascending true* or *descending true*. The important feature of this view is that the context of use determines whether ‘true’ has the content of ‘ascending true’ or the content of ‘descending true’. One problem with the indexicalist option is that users of ‘true’ might not be familiar with ascending truth or descending truth and so might not do what is required for the context of utterance to determine that ‘true’ ends up with one of these contents. It is not even clear how the context of utterance would determine which content ‘true’ would get. Even if there is some intuitive way of letting the context determine its content, there is another problem with this view, which is the topic of the next subsection; roughly, the context of utterance need not have enough information to determine a content for ‘true’.

Option (v): nonindexical contextualism about ‘true’. This is a more contentious view, and some theorists doubt that it is even coherent because it requires denying that the content of an expression determines its extension. In the case of truth, the view treats ‘true’ as univocal and invariant—‘true’ gets assigned a single clause by the presemantics, and this clause disregards the index; so ‘true’ gets the same content regardless of the context of use. A nonindexical contextualist view requires an extra parameter in the points of evaluation for a judge or for standards. This sort of parameter makes sense in cases like predicates of personal taste (for example, ‘fun’) or epistemic modals (for example, ‘might’). Usually, the judge or standard, in addition to any other relevant parameters (for example, worlds or times), together with the proposition expressed by the sentence in a context of use, determines the truth value of the sentence uttered in the context of use. The idea is that an object satisfies the

predicate only if it meets the standard or the judge would say that it does.⁶⁴

Because we have replaced truth with ascending truth and descending truth for the purpose of doing semantics, we altered the semantic theory to have an alethic value parameter, and we altered the post-semantics to output the ascending truth value and the descending truth value of the sentence uttered in a context of use. Because of this change, we have two choices for the nonindexical contextualist theory. The first version adds an alethic standard parameter to the points of evaluation and has the context of utterance determine the alethic standard (either ascending or descending), which gives a reading to the sentence in question just as it does in the assessment-sensitivity option. The postsemantics is still conventional (that is, it outputs ascending truth value in a context of use and descending truth value in a context of use) and would look something like (again, letting u be a context of use, a be a context of assessment, i_u be the index representing u , and i_a be the index representing a):

- (11) A sentence p is *ascending true* at u iff the content assigned to the clause representing p with respect to i_u gets $\mathcal{TM}$ -value 1 at the point of evaluation $\langle w, t, v, s \rangle$, where w , t , and s are the world, the time, and the alethic standard of i_u , and v is the *ascending* alethic value parameter.
- (12) A sentence p is *descending true* at u iff the content assigned to the clause representing p with respect to i_u gets $\mathcal{TM}$ -value 1 at the point of evaluation $\langle w, t, v, s \rangle$, where w , t , and s are the world, the time, and the alethic standard of i_u , and v is the *descending* alethic value parameter.

This version of nonindexical contextualism about truth seems to me to fall prey to a simple objection: those using ‘true’ might not know that it expresses an inconsistent concept and so might not have any idea that the context of utterance is supposed to determine an alethic standard. Moreover, they might not have even heard of ascending truth or descending truth and so would have no idea what an alethic standard is. So it is hard to see how a context of utterance could contribute an alethic standard that would be used by the postsemantic theory.

64. Anil Gupta seems to advocate something like this for inconsistent concepts; see Gupta 1999.

Instead of this version of nonindexical contextualism, there is another that does not have the same problem. It does not rely on the context of utterance to supply an alethic standard, and it is only possible because we have replaced truth in our theory with ascending truth and descending truth. The key is to let the alethic *value* parameter (which is already present in the points of evaluation because we have replaced truth with ascending truth and descending truth) do duty as an alethic *standard* parameter. The points of evaluation have worlds, times, and alethic values—so we have only three slots. The postsemantic theory looks like this:

- (13) A sentence p is *ascending true* at u iff the content assigned to the clause representing p with respect to i_u gets $\mathfrak{tM}$ -value 1 at the point of evaluation $\langle w, t, v \rangle$, where w and t are the world and the time of i_u , v is the *ascending* alethic value parameter.
- (14) A sentence p is *descending true* at u iff the content assigned to the clause representing p with respect to i_u gets $\mathfrak{tM}$ -value 1 at the point of evaluation $\langle w, t, v \rangle$, where w and t are the world and the time of i_u , v is the *descending* alethic value parameter.

The $\mathfrak{tM}$ -value assignments are determined in the following way:

- (i) ‘ p is true’ gets $\mathfrak{tM}$ -value 1 at $\langle w_0, t_0, A \rangle$ iff ‘ p is ascending true’ is ascending true at w_0 and t_0 .
- (ii) ‘ p is true’ gets $\mathfrak{tM}$ -value 1 at $\langle w_0, t_0, D \rangle$ iff ‘ p is descending true’ is descending true at w_0 and t_0 .

Notice that this postsemantics does *not* take the alethic standard to be given by the context of use. Instead, the ascending truth conditions are given by considering a point of evaluation that has the ascending value, where the sentence in question is read by the ascending standard, while the descending truth conditions are given by considering the point of evaluation that has the descending value, where the sentence is read by the descending standard. That is a big difference between this version of nonindexical contextualism and the previous one. Note that this would not work in general for other inconsistent concepts since the postsemantic theory would not be formulated in terms of them. Nevertheless, it offers a very pretty theory of our inconsistent concept of truth without resorting to a radical kind of relativism (that is, assessment-sensitivity). Below, I explain why I prefer the assessment-sensitivity theory over this one; roughly, this theory classifies liar sentences and truth-teller sentences alike, but the assessment-sensitivity view distinguishes between them.

3.4. *The Content Determination Condition*

Two of the alternative theories from the last section—the ambiguity theory (iii) and the indexical theory (iv)—share a common problem: they imply that when someone utters a sentence with an occurrence of ‘true’, the context of utterance determines in part the content expressed by ‘true’ in the sentence uttered. However, because of the way truth predicates are used, there will be many situations where the content of ‘true’ (according to one of these theories) is simply not determined by the information in the context of utterance.

To develop this objection, let us be a bit more precise about the context of utterance by using Robert Stalnaker’s theory of discourse, which focuses on assertion and presupposition. The main idea is that when a speaker makes an assertion in a conversation, the content of the sentence asserted furthers the conversation in a certain way. In particular, the content rules out ways the world might be that were previously live options in the conversation. If the content asserted is accepted by everyone in the conversation, then the potential ways the world might be have been narrowed, and that is one of the central goals of conversation (see Stalnaker 1970, 1974, 1978, 1998, 1999).

To model this idea, assume that we have a conversation consisting of several people. Each person has many beliefs. Stalnaker defines a participant’s *presupposition* as a purportedly shared belief in the conversation. It requires that the participant believes it, the participant believes that everyone else believes it, the participant believes that everyone else believes that everyone else believes it, and so on. Since beliefs are often taken to be attitudes toward propositions, and propositions are often taken to determine a set of possible worlds in which they are true, we can simplify matters by talking about propositions. A participant’s presupposition will divide the class of possible worlds into two—those in which the presupposed proposition is true, and the rest. If a participant’s presupposition is also a presupposition of all the other participants, then it is a *shared presupposition*. The set of shared presuppositions is called the *common ground*—it is what everyone in the conversation agrees on, agrees they agree on, and so on. The crucial notion for Stalnaker’s view is the set of possible worlds in which all the propositions in the common ground are true; call this the *context set*. As the conversation develops, the common ground expands and the context set shrinks. When a participant in the conversation makes an assertion, the proposition asserted should not be entailed by the common ground; that is, it should be false in some

worlds in the context set prior to the assertion.⁶⁵ That way, if everyone in the conversation accepts the assertion, it narrows the context set. Stalnaker's model of conversation has been extremely influential and offers a powerful explanation for a variety of pragmatic phenomena.

In the following passage, Stalnaker (1999, 6) appeals to his views on conversational contexts to formulate a condition on content determination:

It is a substantive claim that the information relevant to determining the content of context-dependent speech acts is presumed to be available to the participants of a conversation—that it is included in the presuppositions of the context—but it is a claim that is motivated by natural assumptions about the kind of action one performs in speaking. It is not unreasonable to suppose that speakers, in speaking, are normally aiming to communicate—at least to have the addressees understand what is being said. Succeeding in this aim requires that the information relevant to determining content be available to the addressee.⁶⁶

Here Stalnaker claims that the information needed to determine the content of an expression (together with the character assigned to it) is part of the common ground.⁶⁷ I agree with this condition, at least insofar as natural language expressions go (one could always make up an expression that violates it, but it is hard to see how such an expression would end up being used by people in actual conversations). However, in order to generate the objection we can be more liberal on the basis for determining the content of expressions—we can say that the information needed to determine content is either part of the common ground or could easily be added to the common ground by, say, querying the participants in the conversation. Moreover, it seems to me that this condition holds not only for indexicals but for ambiguous expressions as well. I refer to it as the *Content Determination Condition*.

Consider the Content Determination Condition in light of the example from section 2.6:

65. Obviously, this feature of Stalnaker's model is an idealization since it would rule out asserting necessary propositions (for example, in mathematical conversations).

66. See also Roberts 2012 [1996], 2010, where she defends the availability thesis: "In order for an utterance to be rationally cooperative in a discourse interaction D, it must be reasonable for the speaker to expect that the addressee can grasp the speaker's intended meaning in so-uttering in D."

67. Since speakers are capable of uttering whatever they want, we should think of this condition as applying only to felicitous utterances—that is, those that are in accord with the norms of communication.

- (E) Every complete sentence in section 2.6 of Scharp's "Truth, the Liar, and Relativism" whose first letter is an 'E' is not true.

Imagine that Cletus asserts (E) in a conversational context with Brandine as his audience because he has it on good authority that all the sentences in question are indeed not true. Assume as well that neither Cletus nor Brandine have access to the relevant section of this essay. It is easy to imagine that Cletus's utterance is felicitous (indeed, we frequently use 'true' and 'false' in situations like this). If, say, option (iv) on which 'true' is an indexical is correct, then (E) as uttered by Cletus expresses either the proposition that every complete sentence in section 2.6 of Scharp's "Truth, the Liar, and Relativism" whose first letter is an 'E' is not descending true *or* the proposition that every complete sentence in section 2.6 of Scharp's "Truth, the Liar, and Relativism" whose first letter is an 'E' is not ascending true. Which is it? Since (E) is among the sentences in question, it turns out that not all of them are not ascending true—(E) in particular is ascending true and not descending true (that is, it is unsafe). However, this information is not in the context of utterance. To use Stalnaker's terminology, the common ground does not include the information about whether the sentences in question are unsafe, nor could this information be ascertained by querying the participants in the conversation. Thus, the conversational context does not determine which proposition (E) expresses; that result would indicate that Cletus's utterance is infelicitous. So, given the Content Determination Condition and empirical unsafety, option (iv) gives us the false prediction that Cletus's utterance is infelicitous. It is easy to see that the same results hold for the ambiguity theory (option (iii)) as well.

We should not pick a theory of truth, or any theory for that matter, according to which content-determining information can outrun what is available in a context of use.⁶⁸ In other words, an adequate theory of truth ought to treat the inconsistency in our concept of truth as a *postsemantic* phenomenon; nonindexical contextualism (version two) and assessment-sensitivity are the only ones that comply.

3.5. *An Example*

To illustrate the two options still on the table, I present a toy language with a presemantic theory, the two semantic theories (that is, one with

68. The only exception I could imagine would be an indexical relativist view on which the context of assessment has a content-determining role; see Weatherson 2009.

$\langle w, t, v \rangle$ points of evaluation and one with $\langle w, t, v, s \rangle$ points of evaluation), and the two postsemantic theories (that is, one with ‘ x is ascending true in context c ’ and ‘ x is descending true in context c ’, and one with ‘ x is ascending true in context of use u from context of assessment a ’ and ‘ x is descending true in context of use u from context of assessment a ’).

3.5.1. Syntax for L

Our language L has several kinds of basic expressions:

- (i) *Individual constants*: ‘I’, ‘Clancy’, ‘insanity pepper’, ‘space coyote’, ‘the sentence’.
- (ii) *One-place predicates*: ‘cook’, ‘true’.
- (iii) *Logical connectives*: ‘ $\sim$ ’, ‘ $\wedge$ ’, ‘ $\vee$ ’, ‘ $\rightarrow$ ’.
- (iv) *Operators*: ‘now’, ‘possibly’.

The following are the formation rules for L’s syntax:

- (i) If α is an individual constant, and γ is a one-place predicate, then ‘ $\gamma(\alpha)$ ’ is a sentence.
- (ii) If ϕ and ψ are sentences, then ‘ $\sim\phi$ ’, ‘ $\phi \wedge \psi$ ’, ‘ $\phi \vee \psi$ ’, and ‘ $\phi \rightarrow \psi$ ’ are sentences.
- (iii) If ϕ is a sentence, then ‘now ϕ ’ and ‘possibly ϕ ’ are sentences.

3.5.2. Semantics for L

A *frame* of L is a seven-tuple $\mathfrak{F} = \langle I, D, W, T, S, V, \mathfrak{I} \rangle$ such that:

- (i) I is a nonempty set of indexes where for all $i \in I$, $i = \langle w, t, d, s, o \rangle$, where $w \in W$, $t \in T$, $d \in D$, $s \in S$, and $o \in D$ (that is, i is a world/time/speaker/alethic standard/object five-tuple).⁶⁹
- (ii) D is a nonempty set (that is, the domain), which includes every sentence of L.
- (iii) W is a nonempty set (that is, the worlds).
- (iv) T is the set of real numbers (that is, the times).
- (v) $S = \{S_A, S_D\}$, where S_A is the ascending alethic standard, and S_D is the descending alethic standard.
- (vi) $V = \{A, D\}$, where A is the ascending alethic value, and D is the descending alethic value.
- (vii) $\mathfrak{I}$ is a function that assigns an intension to each individual constant and predicate other than ‘I’ and ‘the sentence’ (defined below).

69. Remember, an index has information about a context of utterance and a context of assessment; it is distinct from the points of evaluation.

Define the following functions on indexes:

- (i) $w(<w, t, d, s, o>) = w$ (that is, the world of the index).
- (ii) $T(<w, t, d, s, o>) = t$ (that is, the time of the index).
- (iii) $A(<w, t, d, s, o>) = d$ (that is, the agent of the index).
- (iv) $s(<w, t, d, s, o>) = s$ (that is, the aletheic standard of the index).
- (v) $o(<w, t, d, s, o>) = o$ (that is, the salient object of the index).

If α is an individual constant of L other than 'I' or 'the sentence' or a predicate of L , then $\mathcal{I}$ assigns to α an intension $\mathcal{I}_\alpha$ in the following way:

- (i) If α is an individual constant other than 'I' or 'the sentence', then $\mathcal{I}_\alpha$ is a function from $W \times T \times V \times S$ to D such that for all $v, v' \in V$ and $s, s' \in S$, $\mathcal{I}_\alpha(w, t, v, s) = \mathcal{I}_\alpha(w, t, v', s')$; that is, the individual constants other than 'I' and 'the sentence' are assigned a member of the domain for each $\langle \text{world, time, aletheic value, aletheic standard} \rangle$ quadruple, and the function is invariant across aletheic standards and aletheic values.
- (ii) If $\alpha = \text{'cook'}$, then $\mathcal{I}_\alpha$ is a function from $W \times T \times V \times S$ to a 2^D such that for all $v, v' \in V$ and $s, s' \in S$, $\mathcal{I}_\alpha(w, t, v, s) = \mathcal{I}_\alpha(w, t, v', s')$; that is, 'cook' is assigned a subset of the domain for each $\langle \text{world, time, aletheic value, aletheic standard} \rangle$ quadruple that is invariant across aletheic standards and aletheic values.
- (iii) If $\alpha = \text{'true'}$, then $\mathcal{I}_\alpha$ is a function from $W \times T \times V \times S$ to a 2^D ; that is, 'true' is assigned a subset of the domain for each $\langle \text{world, time, aletheic value, aletheic standard} \rangle$ quadruple.

If α is an individual constant or predicate, then the extension of α for frame $\mathcal{F}$, index i , world w , time t , aletheic value v , and aletheic standard s (written as $[[\alpha]]_{\mathcal{F}, i, w, t, v, s}$) is defined in the following way:

- (i) If α is a predicate or an individual constant other than 'I' or 'the sentence', then $[[\alpha]]_{\mathcal{F}, i, w, t, v, s} = \mathcal{I}_\alpha(w, t, v, s)$.
- (ii) $[[I]]_{\mathcal{F}, i, w, t, v, s} = A(i)$
- (iii) $[[\text{the sentence}]]_{\mathcal{F}, i, w, t, v, s} = O(i)$

Define the following extensions for sentences (for frame $\mathcal{F}$, index i , world w , time t , aletheic value v , and aletheic standard s), where α is a constant, γ is a one-place predicate, and ϕ and ψ are sentences:

- (i) $[[\gamma(\alpha)]]_{\mathcal{F}, i, w, t, v, s} = 1$ iff $[[\alpha]]_{\mathcal{F}, i, w, t, v, s} \in [[\gamma]]_{\mathcal{F}, i, w, t, v, s}$
- (ii) $[[\sim\phi]]_{\mathcal{F}, i, w, t, v, s} = 1$ iff $[[\phi]]_{\mathcal{F}, i, w, t, v, s} = 0$
- (iii) $[[\phi \wedge \psi]]_{\mathcal{F}, i, w, t, v, s} = 1$ iff $[[\phi]]_{\mathcal{F}, i, w, t, v, s} = 1$ and $[[\psi]]_{\mathcal{F}, i, w, t, v, s} = 1$
- (iv) $[[\phi \vee \psi]]_{\mathcal{F}, i, w, t, v, s} = 1$ iff $[[\phi]]_{\mathcal{F}, i, w, t, v, s} = 1$ or $[[\psi]]_{\mathcal{F}, i, w, t, v, s} = 1$
- (v) $[[\phi \rightarrow \psi]]_{\mathcal{F}, i, w, t, v, s} = 1$ iff if $[[\phi]]_{\mathcal{F}, i, w, t, v, s} = 1$, then $[[\psi]]_{\mathcal{F}, i, w, t, v, s} = 1$

- (vi) $[[\text{Now } \phi]]_{\mathfrak{F}, i, w, t, v, s} = 1$ iff $[[\phi]]_{\mathfrak{F}, i, w, t, v, s} = 1$
- (vii) $[[\text{Possibly } \phi]]_{\mathfrak{F}, i, w, t, v, s} = 1$ iff for some w' $[[\phi]]_{\mathfrak{F}, i, w', t, v, s} = 1$

Now to define the contents for each sentence, constant, or predicate ζ :

- (i) $\{\zeta\}_{\mathfrak{F}, i} =$ the function from $\langle w, t, v, s \rangle$ to $[[\zeta]]_{\mathfrak{F}, i, w, t, v, s}$.

Contents are functions from points of evaluation to extensions; in the case of sentences, extensions are $\mathfrak{tM}$ -values.

The above definition of a frame says little about how ‘true’ works other than that its $\mathfrak{tM}$ -value might depend on the aletheic value parameter and the aletheic standard parameter—it says nothing about how these parameters function. It will be instructive to consider a particular class of frames that model an example conversation and a particular world and range of times in which the conversation takes place. In this class of frames, the domain includes an insanity pepper, a space coyote, Clancy, and of course all the sentences of L . Every frame in the class has an intension function that assigns (for the world and times in question) the following extensions: Clancy to ‘Clancy’, the insanity pepper to ‘insanity pepper’, the space coyote to ‘space coyote’, and all and only the cooks in the domain to ‘cook’.

In this class of frames, at the world and times in question, the intension of ‘true’ is a function that behaves in the following way. For sentences that do not contain ‘true’, the aletheic standard parameter and aletheic value parameter have no effect. For sentences of L that contain ‘true’, the ascending aletheic standard treats ‘true’ as an ascending truth predicate and the descending standard treats ‘true’ as a descending truth predicate. So it seems like a simple substitution would do the trick. However, because some sentences attribute truth to other sentences that themselves contain truth predicates, we need to be a bit more global in our approach. Let L' be the language that results from substituting ‘ascending true’ for ‘true’ in each sentence of L , and let L'' be the language that results from substituting ‘descending true’ for ‘true’ in each sentence of L . For each sentence ϕ of L , there is a corresponding sentence ϕ' of L' and a sentence ϕ'' of L'' . If ϕ does not contain a truth predicate, then $\phi = \phi' = \phi''$. Now, the ascending standard treats the proposition in question, $\{\phi\}_{\mathfrak{F}, i}$ as if it were $\{\phi'\}_{\mathfrak{F}, i}$, the proposition expressed by the corresponding sentence of L' , and the descending standard treats $\{\phi\}_{\mathfrak{F}, i}$ as if it were $\{\phi''\}_{\mathfrak{F}, i}$, the proposition expressed by the corresponding sentence of L'' . In addition, the intension function for ‘true’ in the class of frames in question might assign different $\mathfrak{tM}$ -values to a

sentence at points of evaluation with different alethic value parameters. At points with the ascending alethic value, a sentence ϕ of L is assigned $\mathfrak{M}$ -value 1 iff the corresponding sentence (that is, ϕ , ϕ' , or ϕ'') is ascending true at the world and time in question, and at points with the descending alethic value, ϕ is assigned $\mathfrak{M}$ -value 1 iff the corresponding sentence is descending true at the world and time in question. In sum, the intension function for 'true' in the class of frames in question obeys the following constraints:

- (i) ϕ gets $\mathfrak{M}$ -value 1 at world w , time t , alethic value A , and alethic standard S_A iff ϕ' is ascending true at w and t .
- (ii) ϕ gets $\mathfrak{M}$ -value 1 at world w , time t , alethic value A , and alethic standard S_D iff ϕ'' is ascending true at w and t .
- (iii) ϕ gets $\mathfrak{M}$ -value 1 at world w , time t , alethic value D , and alethic standard S_A iff ϕ' is descending true at w and t .
- (iv) ϕ gets $\mathfrak{M}$ -value 1 at world w , time t , alethic value D , and alethic standard S_D iff ϕ'' is descending true at w and t .

Thus, even though 'ascending true' and 'descending true' are not part of L , they show up in the semantics to handle the alethic standards and the alethic values.

3.5.3. Presemantics and Postsemantics for L

Now that we have our general semantic theory and our particular class of frames, imagine that Ned and Clancy are having a conversation in English; in their immediate vicinity is a blackboard they can write on. Assume that Clancy is a cook and that Ned is not a cook throughout their conversation.

The Assessment-Sensitivity Option: At a certain point in their conversation Clancy asserts 'I am a cook'. The presemantic theory selects the sentence of L , 'cook(I)', to represent Clancy's sentence and an index $\langle w, t, d, s, o \rangle$ to represent the context of Clancy's utterance (w is the world, t is the time, d is the agent, s is the alethic standard, and o is the salient object—which will be handy when they start talking about the sentence on the blackboard). Note that the alethic standard is picked up from the context of assessment, not the context of utterance, but for ease of exposition, there is just one index that has information from the context of utterance and the context of assessment. Since Clancy's sentence does not have a truth predicate, the alethic standard plays no role at all, and since his sentence is safe, its ascending truth value and descending truth

value are the same. Thus, the semantic theory above assigns 1 to the clause relative to the index and all the world/time/aletheic value/aletheic standard quadruples where Clancy is a cook, and 0 otherwise.

The following is our postsemantic theory:

- (15) A sentence p is *ascending true* in u from a iff the content assigned to the clause representing p with respect to i_u gets $\mathfrak{M}$ -value 1 at the point of evaluation $\langle w, t, v, s \rangle$, where w and t are the world and time of i_u , v is the *ascending* value, and s is the aletheic standard from i_a .
- (16) A sentence p is *descending true* in u from a iff the content assigned to the clause representing p with respect to i_u gets $\mathfrak{M}$ -value 1 at the point of evaluation $\langle w, t, v, s \rangle$, where w and t are the world and time of i_u , v is the *descending* value, and s is the aletheic standard from i_a .

Our postsemantics will interpret 1 as being true in the model and 0 as false in the model. Our postsemantics implies that Clancy's sentence 'I am a cook' is ascending true in the context of use iff Clancy is a cook; it implies that Clancy's sentence 'I am a cook' is descending true in the context of use iff Clancy is a cook.⁷⁰

At another point in their conversation, Clancy writes 'Clancy is a cook' on the blackboard and utters 'the sentence is true'. The presemantic theory selects 'true(the sentence)' to represent Clancy's sentence and an index to represent the context of use. The index, however, also picks up the aletheic standard from the context of assessment, so let us stipulate that the context of assessment is the same as the context of utterance and that Clancy decides on the ascending standard (we will consider the other choice in a moment). In L , 'the sentence' acts like an indexical, and it picks out whatever object is salient in the context—here, it is the sentence written on the blackboard.⁷¹ I stipulated that every sentence of L is in the domain, so there is no problem picking out 'Clancy is a cook'. The clause representing Clancy's sentence 'true(the sentence)' is assigned a proposition, which is then assigned $\mathfrak{M}$ -values at each point of evaluation. The

70. Since 'Clancy is a cook' has no assessment-sensitive vocabulary, the postsemantics does not need to consider contexts of assessment at all. Alternatively, we could say that 'I am a cook' is ascending true in the context of use from any context of assessment iff Clancy is a cook and that 'I am a cook' is descending true in the context of use from any context of assessment iff Clancy is a cook.

71. Again, this feature is an easy way to generate self-reference, but it is not a realistic take on the semantics for locutions like 'the sentence'.

$\mathfrak{tM}$ -value assigned to it at a point $\langle w, t, v, s \rangle$, where $s = S_A$ is the $\mathfrak{tM}$ -value that would be assigned to the proposition expressed by the corresponding sentence of L' at w, t , and v : ascending true (the sentence). That is, the points of evaluation with the ascending alethic standard treat the proposition in question as if Clancy had uttered a sentence containing 'ascending true' instead of 'true'. (Likewise, the points of evaluation with the descending truth standard treat the proposition in question as if Clancy had uttered a sentence containing 'descending true' in place of 'true'. The $\mathfrak{tM}$ -value assigned to it at a point $\langle w, t, v, s \rangle$, where $s = S_D$ is the $\mathfrak{tM}$ -value that would be assigned to the proposition expressed by the corresponding sentence of L'' at w, t , and v . This would be relevant if Clancy had decided on the descending alethic standard instead.) Since "Clancy is a cook" is ascending true' is descending true (and thus ascending true) at the world and time in question, the semantic theory assigns 1 to the proposition at $\langle w_0, t_0, A, S_A \rangle$ and at $\langle w_0, t_0, D, S_A \rangle$, where w_0 and t_0 are the world and time of the index representing Clancy's context of use.⁷² Thus, the postsemantic theory implies that 'the sentence is true' is ascending true in the context of use from the context of assessment and that 'the sentence is true' is descending true in the context of use from the context of assessment. Had Clancy chosen the descending standard instead, the results would have been the same. (The reader will no doubt wonder what motivates this choice—I discuss the issue below.)

Later, Ned writes 'the sentence is not true' on the blackboard. Although we have been considering utterances as verbal performances, let us be a bit more liberal and treat his inscription as an utterance. The presemantic theory selects ' $\sim$ true(the sentence)' to represent Ned's sentence, and the o slot in the index that represents his context picks out ' $\sim$ true(the sentence)' (that is, the clause of the sentence written on the blackboard). In other words, Ned uttered a liar sentence. The proposition assigned to the clause is $\{\sim\text{true}(\text{the sentence})\}_{\mathfrak{F}, i}$; I abbreviate it as ' $\{\phi\}_{\mathfrak{F}, i}$ '. Imagine that it is being assessed by Clancy, who decides to use the descending alethic standard in his context of assessment.

For points $\langle w, t, A, S_A \rangle$ it assigns a $\mathfrak{tM}$ -value based on whether the corresponding proposition $\{\phi'\}_{\mathfrak{F}, i}$ is ascending true at w and t , and at points $\langle w, t, D, S_A \rangle$, it assigns a $\mathfrak{tM}$ -value based on whether the corre-

72. 'Clancy is a cook' is safe, and we have assumed that Clancy is a cook, so 'Clancy is a cook' is descending true (and, of course, ascending true and safe); thus "Clancy is a cook" is ascending true' is descending true (and ascending true).

sponding proposition $\{\phi'\}_{\mathfrak{F}, i}$ is descending true at w and t . For points $\langle w, t, A, S_D \rangle$, it assigns a $\mathfrak{M}$ -value based on whether the corresponding proposition $\{\phi''\}_{\mathfrak{F}, i}$ is ascending true at w and t , and at points $\langle w, t, D, S_D \rangle$, it assigns a $\mathfrak{M}$ -value based on whether the corresponding proposition $\{\phi''\}_{\mathfrak{F}, i}$ is descending true at w and t . Since Clancy chose the descending standard in the context of assessment, we look at L'' , where ‘the sentence’ refers to ‘ $\sim$ descending true(the sentence)’, which is not descending true and is ascending true at w and t . Hence, $[[\text{the sentence}]]_{\mathfrak{F}, i, w, t, A, S_D} \in [[\text{ascending true}]]_{\mathfrak{F}, i, w, t, A, S_D}$, so $[[\sim\text{descending true(the sentence)}]]_{\mathfrak{F}, i, w, t, A, S_D} = 1$.⁷³ Therefore, $\{\phi\}_{\mathfrak{F}, i}$ gets $\mathfrak{M}$ -value 1 at $\langle w, t, A, S_D \rangle$. Moreover, $[[\text{the sentence}]]_{\mathfrak{F}, i, w, t, D, S_D} \notin [[\text{descending true}]]_{\mathfrak{F}, i, w, t, D, S_D}$, so $[[\sim\text{descending true(the sentence)}]]_{\mathfrak{F}, i, w, t, D, S_D} = 0$. Therefore, $\{\phi\}_{\mathfrak{F}, i}$ gets $\mathfrak{M}$ -value 0 at $\langle w, t, D, S_D \rangle$. Remember, the third slot in the points of evaluation contains the alethic value parameter, which determines whether the semantics is assessing the reading of the sentence for ascending truth or descending truth, while the fourth slot contains the alethic standard parameter, which produces a reading of the sentence (it treats ‘true’ in the sentence as either ‘ascending true’ or ‘descending true’).

The postsemantics produces an ascending truth value for the sentence in a context of use from a context of assessment by considering the points of evaluation with the ascending alethic value parameter, and it produces a descending truth value for the sentence in a context of use from a context of assessment by considering the points of evaluation with the descending alethic value parameter. In particular, it yields the following results: Ned’s sentence, ‘the sentence is not true’, is ascending true in Ned’s context of use from Clancy’s context of assessment, and the sentence is not descending true in Ned’s context of use from Clancy’s context of assessment. Since Ned uttered a liar sentence, and Clancy’s context of assessment invokes the descending standard, the semantics reads Ned’s sentence as the descending liar. The ascending alethic value assesses the descending liar for ascending truth—it is ascending true. Then the descending alethic value assesses the descending liar for descending truth—it is not descending true. Had Clancy invoked the ascending standard instead, the results would have been the same.

73. Here I am writing as if ‘ascending true’ and ‘descending true’ have their own clauses in the semantic theory, which would be required for languages L' and L'' . Although I have not specified these clauses explicitly, it should be obvious how it would be done.

The Nonindexical Contextualist Option: The major difference in the nonindexical contextualist theory is that there is no alethic standard slot in the points of evaluation, so they are triples: $\langle w, t, v \rangle$. The alethic value parameter controls the reading of the sentence *and* whether it is being evaluated for ascending truth or descending truth. The postsemantics outputs the ascending truth value of the sentence in the context of use and the descending truth value of the sentence in the context of use.

The first sentence uttered by Clancy, 'I am a cook', is ascending true in the context of use and is descending true in the context of use. The second sentence uttered by Clancy, 'the sentence is true', referring to 'Clancy is a cook', has the same results. It is the third sentence, 'the sentence is not true', that Ned writes on the board, which refers to itself, that we want to investigate. Just as before, the presemantics selects ' $\sim$ true(the sentence)' to represent Ned's sentence, and the o slot in the index that represents his context of utterance picks out ' $\sim$ true(the sentence)'. The semantic theory assigns a character to the clause for Ned's sentence, which together with the index, picks out the proposition that ' $\sim$ true(the sentence)' is not true; call this $\{\phi\}_{\mathfrak{S}, i}$. This proposition is assigned a $\mathfrak{M}$ -value at every point. For points $\langle w, t, A \rangle$, it assigns a $\mathfrak{M}$ -value based on whether the corresponding proposition $\{\phi'\}_{\mathfrak{S}, i}$ is ascending true at w and t , and at points $\langle w, t, D \rangle$, it assigns a $\mathfrak{M}$ -value based on whether the corresponding proposition $\{\phi''\}_{\mathfrak{S}, i}$ is descending true at w and t . In L' , 'the sentence' refers to ' $\sim$ ascending true(the sentence)', which is ascending true and not descending true at w and t . Hence, $[[\text{the sentence}]]_{\mathfrak{S}, i, w, t, v} \in [[\text{ascending true}]]_{\mathfrak{S}, i, w, t, v}$ so $[[\sim\text{ascending true(the sentence)}]]_{\mathfrak{S}, i, w, t, v} = 1$. Therefore, $\{\phi\}_{\mathfrak{S}, i}$ gets $\mathfrak{M}$ -value 1 at $\langle w, t, A \rangle$. On the other hand, in L'' , 'the sentence' refers to ' $\sim$ descending true(the sentence)', which is ascending true and not descending true at w and t . Hence, $[[\text{the sentence}]]_{\mathfrak{S}, i, w, t, v} \notin [[\text{descending true}]]_{\mathfrak{S}, i, w, t, v}$ so $[[\sim\text{descending true(the sentence)}]]_{\mathfrak{S}, i, w, t, v} = 0$. Therefore, $\{\phi\}_{\mathfrak{S}, i}$ gets $\mathfrak{M}$ -value 0 at $\langle w, t, D \rangle$. The postsemantics yields the following results: Ned's sentence, 'the sentence is not true', is ascending true in his context, and his sentence is not descending true in his context. Since the ascending value parameter reads his sentence as the ascending liar, it treats Ned's sentence as ascending true in his context. Since the descending value parameter reads his sentence as the descending liar, it treats Ned's sentence as not descending true in his context. Notice that there are no contexts of assessment, and the reading of the sentence has nothing to do with the context of utterance.

On the nonindexical contextualist option, the reading of Ned's sentence and the evaluation of that reading are controlled by a single slot, so there are only two options—assessing the ascending liar for ascending truth or assessing the descending liar for descending truth. However, on the assessment-sensitivity option, there are two slots, and hence four options—assessing the ascending liar for ascending truth and for descending truth, and assessing the descending liar for ascending truth and for descending truth.

3.6. *Resolving the Paradoxes*

I have not said why I prefer the assessment-sensitivity theory over the nonindexical contextualist theory. There are pros and cons of each. The latter is surely simpler (that is, only one extra parameter), and I would prefer it if the former were not more versatile (that is, two degrees of variability). It is not yet clear whether that additional versatility might come in handy. Note that the two approaches say the same thing about the liar reasoning—it is invalid because it uses (T-In) and (T-Out), both of which have exceptions on these approaches. To justify this claim, we need to take a look at validity.

3.6.1. Validity

Earlier, I said that truth-in-a-model is a mathematical concept, not affected by the paradoxes associated with truth, and the standard way of using truth-in-a-model to define validity remains unaffected as well. The definition of validity is:

(Valid) $\langle \Gamma, \phi \rangle$ is valid iff for every model $\mathfrak{M}$, if all the members of Γ are true-in- $\mathfrak{M}$, then ϕ is true-in- $\mathfrak{M}$.

For our purposes, a model is a point of evaluation in the structure described in the previous section. Thus, an argument $\langle \Gamma, \phi \rangle$ (where ϕ and all the members of Γ are sentences of L) is valid iff for every point of evaluation e , if for all $\gamma \in \Gamma$, $[[\gamma]]_{\mathfrak{F}, i, e} = 1$, then $[[\phi]]_{\mathfrak{F}, i, e} = 1$.⁷⁴

Notice that all the inference rules of classical logic are valid according to (Valid) since every point of evaluation is classical. Notice also that some instances of

(T-In) If ϕ , then $\langle \phi \rangle$ is true, and

(T-Out) If $\langle \phi \rangle$ is true, then ϕ

74. I am assuming a fixed index in this formulation for ease of exposition.

get $\mathfrak{tM}$ -value 0 at some points of evaluation. For example, instances of (T-In) where $\langle\phi\rangle$ is a liar sentence get $\mathfrak{tM}$ -value 0 at points of evaluation with the descending standard, and instances of (T-Out) where $\langle\phi\rangle$ is a liar sentence get $\mathfrak{tM}$ -value 0 at points of evaluation with the ascending standard (in the nonindexical contextualist option, read ‘value’ for ‘standard’ in these claims).

In addition, note that on this inconsistency approach, (T-In) and (T-Out) are constitutive of truth, but they are not ascending true or descending true in general. That is a crucial feature of any inconsistency approach that avoids dialetheism—since the constitutive principles of an inconsistent concept are inconsistent (possibly with respect to some additional assumptions), they cannot all be correct.

3.6.2. The Liar

One point to notice is how these theories deal with paradoxical sentences. For example:

(17) (17) is not true.

A liar argument is below:

1. Assume (17) is true.
2. ‘(17) is not true’ is true [by (T-Sub)].
3. (17) is not true [by (T-Out)].
4. (17) is not true [by reductio].
5. ‘(17) is not true’ is true [by (T-In)].
6. (17) is true [by (T-Sub)].

Our language from the last section, L , cannot express this argument since the only way to refer to its sentences is ‘the sentence’. Still, it would be easy to add ‘(17)’ as a constant, so we could formulate and evaluate this argument (I omit the details).

It is easy to see that the argument is invalid—it fails at two steps. It fails at step 3 because it is not the case that the inference from ‘‘(17) is not true’ is true’ to ‘(17) is not true’ is valid—it fails at points of evaluation with the ascending standard in the assessment-sensitivity view and at points of evaluation with the ascending value on the nonindexical contextualist option. Likewise, the argument fails at step 5 as well, since this step is invalid due to points of evaluation with the descending standard (value).

The other well-known alethic paradoxes are blocked as well for the same reason—neither (T-In) nor (T-Out) are true (-in- $\mathfrak{M}$) at every

point of evaluation. Since the other two familiar alethic paradoxes, Curry's paradox and Yablo's paradox, both require (T-In) and (T-Out), the arguments in them are invalid according to the postsemantic approaches outlined here.

3.6.3. Truth Tellers

Truth tellers are not paradoxical, but many approaches assign to them the same status as paradoxical sentences (for example, Kripke's strong Kleene minimal fixed point, and Field's theory). Other things being equal, it would be good for a theory of truth to treat paradoxical sentences like the liar and merely odd ones like the truth teller differently. Thinking about how truth tellers are handled by the nonindexical contextualist and assessment-sensitivity approaches provides us with additional data that are relevant in deciding between them. The truth teller is:

(18) (18) is true.

Again, L cannot express this sentence, but it is easy to add '(18)' to L so that it can (I omit the details).

Nonindexical contextualism has only one slot for the alethic value, which serves two purposes: a reading of (18) and the entry in the alethic conditions for (18). According to ADT, the ascending truth teller and the descending truth teller are safe; the descending truth teller is not ascending true, and the ascending truth teller is descending true.⁷⁵ Thus, the nonindexical contextualist approach has the following consequences for context *u*: (18) is ascending true in *u*, and (18) is not descending true in *u*. That is, it has the same alethic conditions as the liar.

The assessment-sensitivity view differs on the truth teller. Assessment-sensitivity semantics has two slots—one for alethic standards, which controls the reading of (18), and the other for alethic value, which dictates which alethic condition is relevant. The assessment-sensitivity approach implies that (as used in context *u*) (18) is descending true in *u* from contexts with the ascending standard, and (18) is not ascending true in *u* from contexts with the descending standard. That is significantly different from the status the liar has. Recall that (17) is not descending true in the context of use from contexts with the ascending standard, and (17) is ascending true in the context of use from contexts with the descending standard. Therefore, nonindexical contextualism

75. These results are nontrivial, but I do not have the space to justify them; see Scharp 2011.

cannot distinguish between paradoxical sentences like the liar and merely odd sentences like the truth teller, but the assessment-sensitivity approach can. That is a big point in favor of assessment-sensitivity, and for this reason, I endorse the assessment-sensitivity approach.

3.7. *Problems for Semantic Relativism*

For those well versed in the literature on semantic relativism, probably very little of the discussion of the descriptive theory of truth has been familiar. The way alethic standards work in the theory described above differs from the way most standards work in semantic relativist theories, and the arguments given for the theories above differ from the kind of evidence usually marshaled in favor of semantic relativist treatments. In this section, I explain these differences and consider some common objections to semantic relativism.

The literature on semantic relativism is large and growing every week, it seems. Theorists have argued for some form of semantic relativism for many kinds of linguistic expressions, including predicates of personal taste (see Kölbel [2002, 2003]; MacFarlane [2005a, 2007a, 2008, 2011b, forthcoming]; Egan [2010]; Lasersohn [2005, 2008, 2009]; and Einheuser [2008]) and epistemic modals (see Kölbel [2002]; Egan, Hawthorne, and Weatherson [2005]; Egan [2007]; Stephenson [2007]; MacFarlane [2011a, forthcoming]; Einheuser [2008]). Most of the semantic relativists argue that their theories capture the linguistic data better than the alternatives. The linguistic data include: (i) the surface grammar of the expressions in question, (ii) the ways in which speakers take one another to agree on certain points in certain situations and disagree on certain points in other situations, (iii) the circumstances under which agents treat one another as having said the same thing, (iv) the ways in which speakers treat themselves and one another as authoritative on certain issues, and (v) the ways in which speakers retract claims in light of certain challenges. For example, one might think that if one person asserts that stewed rhubarb is tasty and another asserts that it is not, then they disagree, but neither of them is guilty of some cognitive fault or shortcoming. Instead, this might be a case of faultless disagreement. Some nonindexical contextualists and assessment-sensitivity theorists have argued that their views offer the best explanation of faultless disagreement (see Kölbel [2002, 2003]; Lasersohn [2005]; Recanati [2007]; and MacFarlane [2005a, 2007a, forthcoming]). In addition, some assessment-sensitivity theorists argue that their view explains speakers'

tendency to retract previous claims better than nonindexical contextualism; for example, if Millhouse at age ten asserts that Squishees are tasty but then at age twenty denies that they are tasty, he might say that his age-ten utterance was mistaken. It is difficult for nonindexical contextualism to explain this behavior since it entails that the sentence Millhouse uttered at age ten is true in that context of use. However, the assessment-sensitivity theorist can say that the sentence Millhouse uttered at age ten is true in the age-ten context of use from the age-ten context of assessment, but the sentence he uttered at age ten is false in the age-ten context of use from the age-twenty context of assessment.⁷⁶

In the case of truth, almost all speakers are ignorant of the fact that truth is an inconsistent concept. Thus, speakers use it as if it were consistent; hence the above kinds of linguistic data with respect to truth are not decisive. Since speakers are ignorant of truth's inconsistency, they are bound to make mistakes with it. We already know that we do not see faultless disagreement or retraction data in the case of truth because speakers are unaware that truth is inconsistent, and hence they are unaware that 'true' is assessment sensitive. Remember: it need not be the case that simply possessing an inconsistent concept is enough for its possessor to come to know that it is inconsistent. If words expressing inconsistent concepts are assessment sensitive, then simply being competent with a word is not enough to come to know that it is assessment sensitive.

Notice, however, that I have not argued for an assessment-sensitivity view in the familiar way. Instead, once one accepts an inconsistency approach to the alethic paradoxes, one must choose presemantic, semantic, and postsemantic theories for truth on the basis of more general principles about language use, like the Content Determination Condition defended above. Indeed, this condition gives us good reason to think that if truth is an inconsistent concept, then only a postsemantic approach to the alethic paradoxes is acceptable. The reason is that truth is inconsistent by virtue of the empirical context in which it is used—that is, the empirical context in which it is used is one where rational agents reason more or less classically and speak natural languages that have the capacity to represent their own syntax.⁷⁷ These are contingent features

76. See MacFarlane 2005a, 2007a, forthcoming, for discussion.

77. That is, deviations from classical logic are minor (for example, intuitionistic logic or the logic of relevant implication) compared to what is required to avoid the paradoxes (for example, LP or K₃). See Field 2008 for a discussion of the logics used by nonclassical approaches to the paradoxes.

of natural languages, and in these environments, any concept that has (T-In) and (T-Out) as constitutive principles is inconsistent. Had our linguistic practices been different, our concept of truth might not have been inconsistent. Therefore, 'true' is assessment sensitive not entirely because of the ways in which speakers use it. It is assessment sensitive because of the ways in which speakers use it together with the logical and syntactic environment in which it is used. Speakers can be, and often are, ignorant of the fact that this environment is hostile to a concept with (T-In) and (T-Out) as constitutive principles. Thus, speakers are ignorant of the fact that 'true' is assessment sensitive. Moreover, a view that introduces contexts of assessment is much more suited to a situation where one group of people with a more advanced conceptual repertoire (for example, theorists interested in the semantics for natural language who possess the concepts of ascending and descending truth) are trying to explain the behavior of another group of people with a less advanced conceptual repertoire (for example, everyday speakers of English who use truth but have never heard of ascending truth or descending truth) because the more advanced concepts play a role only in the contexts of assessment. With an assessment-sensitivity theory, there is no expectation that users of the concept of truth have any familiarity with the replacement concepts.

Because of the Content Determination Condition, it is unacceptable for us to use a presemantic approach (for example, 'true' is ambiguous) or a semantic approach (for example, 'true' is an indexical) because these would imply that speakers and hearers are ignorant of the propositions expressed by sentences containing 'true'. In other words, it cannot be that linguistic expressions are ambiguous or indexical by virtue of the environment in which they are used. However, a linguistic expression can be assessment sensitive by virtue of the environment in which it is used if the concept expressed by that expression is inconsistent in that environment. That, I claim, is exactly the case with truth.

Another major difference between my use of semantic relativism and the use to which it is more often put is in the standards that serve as the extra parameter in points of evaluation. The difference is that standards are usually ways of modifying or narrowing the content of the expression in question, whereas, in the case of truth, they provide an alternative reading of the truth predicate. For example, if one takes 'fun' to be assessment sensitive or nonindexical contextual, then the points of evaluation have an extra slot for an enjoyment standard. A standard of enjoyment often just specifies that in which a particular person

finds enjoyment. The standard of enjoyment does not reinterpret the word ‘fun’; rather, since what is fun seems to be subjective, it provides the extra bit of information that allows one to assign an extension to ‘fun’. On the other hand, in the case of truth, the alethic standards do *not* encode what a particular person takes to be true, and the motivation for semantic relativist treatment is *not* some antecedent intuition that truth is somehow subjective. Instead, there are only two possible alethic standards, and they reinterpret the word ‘true’ so that the semantic theory can assign it an extension without falling into contradiction—one reads ‘true’ as ‘ascending true’ and the other reads it as ‘descending true’.⁷⁸ There is no reason to think that a person will adopt one of these alethic standards but not the other—the alethic standards are not personal. Instead, someone who knows that truth predicates are assessment sensitive can switch back and forth between the two in order to get a better grip on the semantic features of some sentence with an occurrence of ‘true’. Unlike standards of enjoyment, which are often thrust upon us by our responses to external stimuli, an alethic standard is chosen by an interpreter based on whether it makes more sense to consider the truth claim in question as an ascending truth claim or a descending truth claim at that moment.

By treating ‘true’ as assessment sensitive, those of us who both recognize that it expresses an inconsistent concept and possess the proper replacement concepts can interpret people who use ‘true’ in a consistent way—we can understand their assertions, assess their claims, and evaluate their arguments, all without contradicting ourselves, giving up any cherished logical principles, or treating them as if they do not understand the content of ‘true’. In short, the assessment-sensitive theory of truth uses the semantic relativist’s formal machinery but interprets it in a new way that satisfies one demand of an inconsistency approach to the alethic paradoxes.

A standard criticism of semantic relativism is that contextualism does a better job of explaining the data (see Glanzberg [2007, 2011]; Cappelen [2008]; Cappelen and Hawthorne [2009, 2011a, 2011b]; von Fintel and Gilles [2008]; and Schaffer [2011]). However, this sort of objection has no bite in the case of truth. Very few possessors of the concept of truth are aware that it is an inconsistent concept, and very few users of truth predicates are aware that they are assessment sensitive. As such, very few users of truth predicates treat them as one would treat an

78. This feature is not unprecedented—the supervaluation semantics considered in section 3.3 is similar.

assessment-sensitive expression or as one would treat an indexical. Once one accepts an inconsistency approach to the paradoxes, one has to decide between contextualism and semantic relativism on other considerations. I have argued that contextualism is not a good candidate for 'true' because of the combination of empirical unsafety and the Content Determination Condition. Semantic relativism, however, does not run afoul of the Content Determination Condition.

One might, however, worry that the assessment-sensitivity theory does have a similar problem. After all, if truth is assessment sensitive, but very few people know that it is, then very few people will know to adopt alethic standards in contexts of assessment. How can the assessment-sensitivity theory accurately describe and explain the facts about how truth predicates function in our natural languages when almost no one ever adopts an alethic standard? Wouldn't we need to educate all English speakers about truth's inconsistency and the replacement concepts before the assessment-sensitivity theory could be expected to deliver the right results?

The answers to these questions and the key to replying to this objection come by answering another question: who are the consumers of approaches to the alethic paradoxes? They are for people who are bothered by the problems posed by the alethic paradoxes—problems like: (i) we can derive intuitively absurd conclusions from intuitively obvious assumptions via intuitively obvious inferences, and (ii) one arrives at inconsistent results when one tries to use a standard semantic theory to explain semantic features of truth predicates. These are the people to whom I am recommending the assessment-sensitivity theory of truth, and these are the people who, if they want solutions to their problems, will need to get up to speed on the concepts of ascending truth and descending truth so that they can adopt alethic standards in contexts of assessment. People who are ignorant of the alethic paradoxes have no need for, or interest in, possessing these concepts or adopting these standards. The theory is for the theorists—those who want to make sense of those who use our inconsistent concept of truth. It is only those who use the theory that adopt alethic standards and so need to possess the concepts of ascending truth and descending truth. That is exactly the major benefit of the assessment-sensitivity view. On the ambiguity view or the indexical view, speakers choose between *ascending true* and *descending true* as the content of 'true', which is exceedingly implausible since most speakers do not possess these concepts. On the assessment-sensitivity view, speakers just utter sentences containing 'true', and 'true' has an

invariant content. Hearers attribute contents to the claims made by speakers in the same intuitive way. The contexts of assessment and the alethic standards are for theorists who want to attribute ascending truth conditions and descending truth conditions to sentences containing 'true'. There is no reason to think that ordinary speakers or hearers would have any interest in such a thing. The only people who should care about ascending truth conditions and descending truth conditions are those of us who want to solve the problems posed by the alethic paradoxes.

Consider another objection: Crispin Wright argues that semantic relativist postsemantic theories are appropriate only for propositions that are nonrepresentational because these theories allow for faultless disagreement. For example, an assessment-sensitivity theory of knowledge would be acceptable, according to Wright, only if there is nothing for knowledge attributions to represent. Presumably, many of us have the intuition that at least some knowledge attributions represent certain mental states of certain people. Thus, Wright's (2008) point, if correct, would pose a problem for these accounts of knowledge.⁷⁹ Wright's point, if it is correct, does not affect the assessment-sensitivity theory of truth since I do not think there is any property of truth to be represented by our word 'true' and our concept of truth. Instead, there are two properties, being ascending true and being descending true, and anyone who thinks that there is a property of being true is confused. Thus, the claim that 'true' does not represent something (that is, the property of being true) is one I endorse. The assessment-sensitivity view of truth does allow for limited faultless disagreement (in cases where the difference between ascending truth and descending truth is negligible). Moreover, I have argued that, together, ascending truth and descending truth can give an adequate account of the representational aspect of content insofar as they play an explanatory role in the revised semantic and postsemantic theories offered above. Thus, Wright's worry is not a problem for my proposal.

In sum, the assessment-sensitivity approach is compatible with the Content Determination Condition and is fit for inclusion in an inconsistency approach to the alethic paradoxes; with the exception of the nonindexical contextualist approach, the others studied here are not. It makes sense to say that a linguistic expression can turn out to be

79. See MacFarlane 2005b for an assessment-sensitivity theory of knowledge.

assessment sensitive by virtue of the way it is used and the environment in which it is used (which might be unknown or unrecognized by those who use it), whereas it does not make sense to say this about indexicality or ambiguity. The standard objections to assessment-sensitivity are inapplicable to the case of truth since these objections presuppose that speakers would be aware of the fact that their expressions are assessment sensitive. At some point after the aletheic revolution, when those for whom it is relevant know to distinguish between ascending truth and descending truth, we will have more data on how these people use 'true', and we might need to reassess whether the standard objections have any bite.

3.8. Discussion

Now that we have a sketch of the right kind of inconsistency approach, I would like to briefly discuss several issues that naturally arise in connection with this kind of proposal.

3.8.1. Expressive Role

I have already mentioned that we use 'true' as a device of endorsement and as a device of rejection. Earlier, I argued that the replacement concepts can serve these roles. Here, I consider the extent to which 'true' as described by the assessment-sensitivity theory plays these roles.

According to the inconsistency approach, (T-In) and (T-Out) are constitutive of the concept of truth, which is the reason it is an inconsistent concept. Since (T-In) and (T-Out) explain these expressive roles, the inconsistency approach can make sense of why those who use truth predicates *take them* to serve these expressive roles. Participants in linguistic practices take (T-In) and (T-Out) to be constitutive of truth, which leads them to think these principles are true, which leads them to use truth predicates as devices of endorsement and as devices of generalization. For example, in a conversation between Sherri and Terri, Sherri asserts 'Frink's theory is true'. Both Sherri and Terri take (T-In) and (T-Out) to be constitutive of 'true', so they both take Sherri to have committed herself to the claims that constitute Frink's theory, regardless of the content of those claims or whether Sherri or Terri knows what Frink's theory says. That is an example of 'true' being used as a device of endorsement, and the inconsistency approach explains why people use it that way.

However, according to the assessment-sensitivity theory, (T-In) and (T-Out) have exceptions—there are sentences *p* such that '*p* is true' does not follow from *p*, and there are sentences *q* that do not follow

from 'q is true' (it turns out that exceptions to one will also be exceptions to the other). Thus, if one calls one of these sentences true, then one has not thereby endorsed that sentence. For example, assume that one of the sentences that make up Frink's theory is an exception (perhaps it is like the example of an empirically paradoxical sentence from section 2.6). Call it 'p'. When Sherri asserts that Frink's theory is true, she commits herself to all the sentences of Frink's theory that are not exceptions, but she does not commit herself to p since p does not follow from 'p is true'. Even though Sherri and Terri both assume that she has committed herself to p, they are wrong. Thus, the inconsistency approach predicts that people will use truth predicates as devices of endorsement because of its constitutive principles, but assessment-sensitivity theory predicts that there will be some mistakes about these uses since users do not realize that these constitutive principles are inconsistent (given the existence of paradoxical sentences).

Why is this not just as bad as traditional approaches that deny (T-In) or (T-Out)? Easy—these approaches presuppose that truth is a consistent concept, and as descriptive theories, they make certain predictions about how speakers should use 'true'; but if we find that speakers do not use 'true' in this way, it is difficult to explain away the discrepancy by saying that speakers are just making mistakes. On the other hand, if truth is an inconsistent concept, then one should expect to find those who are unaware of this fact make some mistakes when using it. What kind of mistakes? I have argued that it is wrong to see them as making mistakes about the *content* of 'true' — 'true' has exactly the constitutive principles it seems to have. Instead, their mistake is using a concept that, unbeknownst to them, is inconsistent. The assessment-sensitivity approach does *not* predict that people will use 'true' in accordance with the semantics given above. Instead, it entails that those who know that truth is an inconsistent concept should use it that way, and it instructs those of us who know that truth is an inconsistent concept to interpret the rest of the users of 'true' in that way.

3.8.2. Revenge

The revenge paradox phenomenon is the nemesis of virtually everyone who sets out to solve the liar. Revenge paradoxes are almost impossible to avoid, and they thwart most attempts to say something consistent (or at least nontrivial) about the alethic paradoxes. However, the inconsistency approach outlined here does not give rise to any revenge paradoxes.

How does it avoid them? Consider a sentence that might seem to give rise to a revenge paradox:

(19) (19) is either false or paradoxical.

One might try to argue that (19) causes a problem for the assessment-sensitivity approach in the following way:

1. Assume (19) is true.
2. '(19) is either false or paradoxical' is true.
3. (19) is either false or paradoxical.
4. Assume (19) is either false or paradoxical.
5. '(19) is either false or paradoxical' is true.
6. (19) is true.

If this argument were valid, it would be a problem, but it is not. The assessment-sensitivity approach is fully classical, so the logical inferences in this argument are all fine. However, since (19) is paradoxical (that is, ascending true and not descending true from any context of assessment), the move from step 2 to step 3 is invalid, and the move from step 4 to step 5 is invalid. The upshot is: the assessment-sensitivity approach implies that (19) is paradoxical, but it does not imply that '(19) is paradoxical' is true. So, of course, the assessment-sensitivity approach has consequences that it deems untrue. Is this a problem? It would be a problem if the assessment-sensitivity approach implied that truth is a consistent concept, but it does not. Instead, the assessment-sensitivity approach implies that truth is an inconsistent concept that cannot be legitimately applied in every circumstance. Indeed, the theory of ascending truth and descending truth specifies exactly where truth can be used without running into problems (that is, when the difference between ascending truth and descending truth is negligible). Accordingly, the assessment-sensitivity approach is outside the legitimate scope of truth—in applying truth to it, one gets odd results. Consider the analogy. Even though the concept of mass is inconsistent, it is fine to use it in certain situations (that is, when the difference between relativistic mass and proper mass is negligible). However, if one tries to use mass outside these situations, say, in calibrating the atomic clocks in GPS satellites, it will not provide accurate predictions. Instead, in these circumstances, one needs to use the replacements.

In the case of truth, one can reasonably ask whether the assessment-sensitivity approach is ascending true and whether it is descending true. All the central principles of the assessment-sensitivity approach are descending true. However, descending truth is not preserved under

logical consequence, so it could turn out that the assessment-sensitivity approach has some consequences that are not descending true. They would, of course, be ascending true. I have been unable to identify any consequences of the assessment-sensitivity approach that have this feature, but I have not been able to rule it out either. If it does have these kinds of consequences, then it would be in the same boat as the ADT. Either way, there are not any revenge paradoxes here.

One consequence of the dearth of revenge paradoxes is that the inconsistency approach presented here does not require any expressive impoverishment in the languages to which it applies. It needs no object language/metalanguage distinction, and it works perfectly well for languages that have the capacity to express the theory of ascending and descending truth (ADT) and the assessment-sensitivity theory of 'true'.⁸⁰

3.8.3. Other Approaches to the Paradoxes

Consider the options for addressing the alethic paradoxes in a bit more detail. The innovation in this subsection is that they are split into *philosophical approaches*, which discuss some feature of truth predicates of natural languages relevant to resolving the paradoxes, and *logical approaches*, which focus on artificial languages and propose principles that truth predicates obey and logics for languages containing truth predicates. Philosophical approaches to the alethic paradoxes do two things: they tell us something about the truth predicate that is relevant to solving the alethic paradoxes, and they tell us something about the paradoxical truth bearers and the paradoxical reasoning. Logical approaches specify principles that truth predicates obey and logics that are compatible with these principles. The theories of truth offered by logical approaches apply to certain artificial languages, and these theorists use techniques from mathematical logic to investigate the properties of these theories and to prove things about them (for example, consistency relative to a background mathematical theory). Although this distinction is not common, I find it a crucial aid to understanding the range of alternatives since many approaches appeal to mathematical structures, each of which can be (and often is) given different philosophical interpretations by different theorists.⁸¹

80. See Scharp 2013, forthcoming, for details.

81. Field (2008) explicitly formulates most of the logical approaches described below, but he is not as careful about the philosophical approaches and does not consider the space of combinations of the two.

Philosophical approaches to the aletheic paradoxes include:⁸²

1. *Meaningfulness*: paradoxical sentences are meaningless (see Ushenko [1937]; Grover [1977]; Brandom [1994]; Sorensen [2001]; Armour-Garb [2001]; and Beall [2001]).
2. *Ambiguity*: truth predicates are ambiguous (that is, they have different meanings on different occasions of use), and the paradoxical reasoning is invalid by equivocation (see Parsons [1974]; Kripke [1975]; and Williamson [2000]).
3. *Contextual*: truth predicates (or sentences that contain them) are context dependent (that is, they express different contents on different occasions of use), and the paradoxical reasoning is invalid because it violates the logic of indexicals (see Thomason [1976]; Parsons [1974]; Burge [1979]; Barwise and Etchemendy [1987]; Gaifman [1992]; Simmons [1993]; Glanzberg [2004]; and Shapiro [2006]).
4. *Circularity*: truth predicates express circularly defined concepts, and the paradoxical reasoning is invalid because it violates the logic of circularly defined concepts (see Gupta and Belnap [1993]).
5. *Indeterminacy*: truth predicates exhibit indeterminacy, and the paradoxical reasoning is unsound because it violates the logic appropriate for indeterminacy (see van Fraassen [1968]; Kripke [1975]; McGee [1991]; Soames [1999]; Maudlin [2004]; and Field [2008]).
6. *Inconsistency*: truth predicates express inconsistent concepts (or render languages of which they are constituents inconsistent), and either: (i) the paradoxical reasoning is sound because all the principles involved are constitutive or (ii) the paradoxical reasoning is unsound because one of the principles involved is constitutive but false (or invalid).⁸³

Logical approaches to the aletheic paradoxes are classified by the combination of aletheic principles and logical principles compatible with the theory of truth offered. The aletheic principles are:

82. These categories are intended to break up the literature in a natural way, but they should not be taken to be exclusive or exhaustive.

83. For dialetheic examples, see Priest 1979, 2006a; and Beall 2009. For the others, see Chihara 1979; Yablo 1993a, 1993b; Eklund 2002a; Patterson 2006; and Scharp 2007, 2008, 2011, 2013, forthcoming.

- (T-In) If p , then $\langle p \rangle$ is true.
- (T-Out) If $\langle p \rangle$ is true, then p .
- (T-Intro) $p \vdash \langle p \rangle$ is true.
- (T-Elim) $\langle p \rangle$ is true $\vdash p$.

The differences between these principles is crucial. The *conditional* principles, (T-In) and (T-Out), are the strongest. The *inferential* principles, (T-Intro) and (T-Elim), are weaker than the conditional principles (although they are equivalent in logical systems for which one can prove a deduction theorem—that is, those that allow conditional proof). The logics are harder to describe, but there are five main options:

- (i) *Classical logic*, which has all the standard introduction and elimination rules for each connective plus the hypothetical rules.
- (ii) *Weakly classical logic*, which has all the inference rules of classical logic but no meta-rules.⁸⁴
- (iii) *Paracomplete logic*, on which the law of excluded middle and some classical rules are invalid.⁸⁵
- (iv) *Paraconsistent logic*, on which the rule *ex falso quod libet* and some other classical rules are invalid.⁸⁶
- (v) *Substructural logic*, on which some of the structural rules governing derivability are invalid.⁸⁷

When we map out the compatible combinations of alethic principles and logics, we get the following seven options for logical approaches to the alethic paradoxes:

- 1. *Classical Glut*: classical logic and (T-In).⁸⁸
- 2. *Classical Gap*: classical logic and (T-Out) (Feferman [1982]; Maudlin [2004]).
- 3. *Classical Symmetric*: classical logic, neither (T-Intro) nor (T-Elim) (McGee [1991]).

84. That is, weakly classical logics do not validate conditional proof, reductio, or reasoning by cases. See Field 2008, chaps. 10–12, for discussion.

85. Paracomplete logics (for example, K_3) validate *ex falso* and double negation elimination but not conditional proof, reductio, or contraction.

86. Paraconsistent logics (for example, LP) validate excluded middle and double negation elimination but not disjunctive syllogism, *ex falso*, or contraction.

87. See Restall 1994, 2000, and Beall and Ripley, forthcoming, for background on substructural logics.

88. I am unaware of any classical glut theorists; see Field 2008 for discussion.

4. *Weak Classical*: weakly classical logic, (T-Intro), and (T-Elim) (Friedman and Sheard [1987]; Gupta and Belnap [1993]; and Yablo [1993a and b]).
5. *Paracomplete*: paracomplete logic, (T-In), and (T-Out) (Kripke [1975]; Soames [1999]; and Field [2008]).
6. *Paraconsistent*: paraconsistent logic, (T-In), and (T-Out) (Priest [2006a, 2006b]; Beall [2009]).
7. *Substructural*: substructural logic, (T-In), and (T-Out) (Tennant [1997]; Weir [2005]; Zardini [2011]; Ripley [2012]).

Notice that the first two options permit only one (primary) alethic principle, the third has no (primary) alethic principles,⁸⁹ and the other four have two principles—the weakly classical approaches have the moderate-strength alethic principles, and the paracomplete, paraconsistent, and substructural approaches have the strongest alethic principles. Thus, as one might expect, the strength of the logic and the strength of the alethic principles vary inversely.

The obvious question now is: what are the fruitful combinations of philosophical approaches and logical approaches to the paradoxes? Describing each combination would be tedious and not add much to this essay. Instead, I have depicted them in figure 4, which contains a diagram showing philosophical approaches on one side and logical approaches on the other, with important connections between them and the theorists who work on them.

There are several trends to notice. First, the meaningfulness philosophical approach has no connections to the logical approaches—that is because it tries to find some problem with paradoxical truth bearers. Approaches like these do not need logical approaches to handle paradoxical sentences because they imply that there is no problem to be handled.

Second, a single philosophical approach might be paired with distinct logical approaches. For example, the indeterminacy approach is championed by Hartry Field, Tim Maudlin, and Vann McGee. Field pairs it with a paracomplete approach, McGee (1991) with a classical

89. However, some classical symmetric theorists accept very weak principles, for example:

(T-Enter) If $\vdash p$, then $\vdash \langle p \rangle$ is true.
 (T-Exit) If $\vdash \langle p \rangle$ is true, then $\vdash p$.

See McGee 1991 for an example.

Philosophical and Logical Approaches

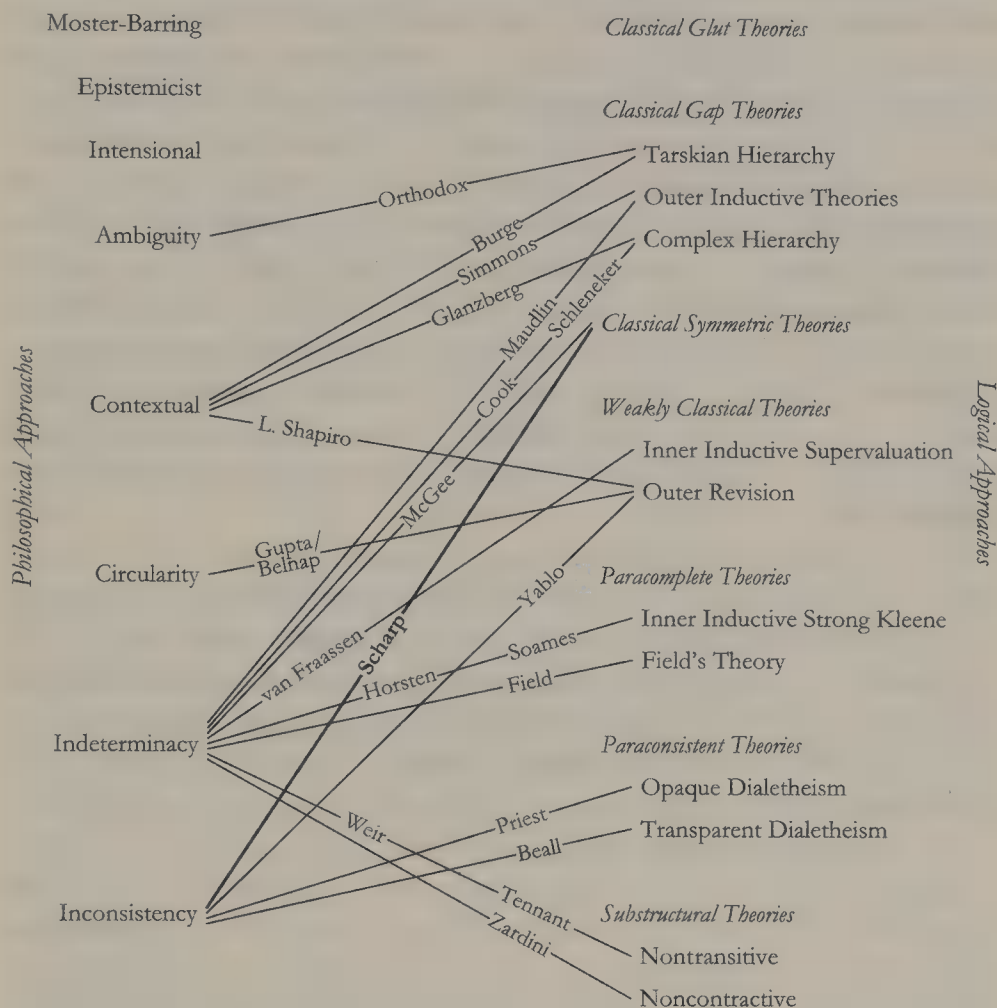


Figure 4.

symmetric approach, and Maudlin (2004) with a classical gap approach.⁹⁰ In each of these combinations, the notion of indeterminacy is interpreted differently. The same point holds of contextual views as well.

Third, a single logical approach might be paired with distinct philosophical approaches. For example, the orthodox approach pairs a Tarskian hierarchy (a classical gap approach) with an ambiguity approach—it interprets ‘true’ as ambiguous so that it can have the meaning of any one of the predicates in the Tarskian hierarchy. Tyler Burge (1979, 1982a,

90. In saying that Maudlin’s approach is classical, I am endorsing Field’s interpretation of it; see Field 2006b.

1982b) uses the same logical approach (that is, the Tarskian hierarchy), but he pairs it with a context-dependence philosophical view. He claims that ‘true’ is an indexical that has an invariant meaning (that is, character) and variable content; in any given context, ‘true’ can have the content of any of the predicates in the Tarskian hierarchy. These are very different interpretations (that is, ambiguity and indexicality) of the same formal structure (that is, the Tarskian hierarchy). Another example is the revision theory, which is a weakly classical (logical) approach and was initially designed by Gupta and Belnap (1993) to be paired with a circularity (philosophical) approach. However, Stephen Yablo (1993a, 1993b) pairs the revision theory with an inconsistency (philosophical) approach, while Lionel Shapiro (2006) suggests that the context-dependence (philosophical) approach is a better fit for revision theories. Again we have two very different philosophical interpretations of the same mathematical structure.

In this essay, I offer the combination of a novel inconsistency (philosophical) approach, which takes words that express inconsistent concepts to be assessment sensitive, and a symmetric classical (logical) approach. Notice that this combination is the first to pair an inconsistency philosophical approach with any kind of classical logical approach. As such, it provides for the first time a classically consistent theory of our inconsistent concept of truth.

In a essay of this length, it would be impossible to compare and contrast the assessment-sensitivity approach offered here with all the others listed in figure 4. Nevertheless, I can summarize my reasons for thinking that it is superior to its competitors. First, as discussed in the introduction, all those who endorse traditional approaches, which include all philosophical approaches besides inconsistency, end up flatly denying principles that are clearly constitutive of either truth or our logical concepts, depending on which logical approach they adopt. Only an inconsistency approach can avoid this problem. Moreover, dialetheic-inconsistency approaches have the same problem—they are for all intents and purposes traditional approaches. Only a nondialetheic-inconsistency approach can avoid this problem, but of those who endorse these approaches (for example, Chihara and Eklund), only Yablo has adopted a logical approach to the paradoxes. His theory has not received much attention, but it implies that the truth value of paradoxical sentences changes continuously, which seems rather implausible; it also is not compatible with classical logic (it is a weakly classical approach). So the assessment-sensitivity theory presented here is the first to combine a non-

dialetheic-inconsistency approach with a classical logical approach. This is not just filling in a slot in logical space. If what I have said is right, then it is the only approach to the alethic paradoxes that does not do violence to our conceptual principles.

Second, the assessment-sensitivity approach is compatible with classical logic and requires no expressive limitations on the languages to which it applies. It is impossible to find this pair of features in any other combination approach to the alethic paradoxes that also entails that (T-In) and (T-Out) are constitutive of our concept of truth. Moreover, the combination of assessment-sensitivity as an inconsistency approach and classical symmetric as a logical approach completely avoids the revenge paradoxes that have been the most difficult problem facing anyone who tries to say something about the liar and its ilk.

Finally, and most importantly, we have had good reason to think that truth is an inconsistent concept for a long time. We also have a precedent for dealing with inconsistent concepts (as illustrated by the case of mass). The approach offered here simply does for truth what we have always done when we find out that one of our useful concepts is inconsistent—replace it with consistent concepts that do its job in certain situations and figure out how to interpret those who continue to use the inconsistent concept in question. On this line of thought, truth joins mass and a host of other prominent concepts in being replaced for certain purposes due to inconsistency.

References

- Antony, Louise. 1997. "Meaning and Semantic Knowledge." *Aristotelian Society Supplementary Volume* 71: 177–207.
- Armour-Garb, Bradley. 2001. "Deflationism and the Meaningless Strategy." *Analysis* 61: 280–89.
- Asher, Nicholas, Josh Dever, and Chris Pappas. 2009. "Supervaluations Debugged." *Mind* 118: 901–33.
- Atlas, Jay David. 1989. *Philosophy without Ambiguity: A Logico-Linguistic Essay*. Oxford: Oxford University Press.
- Barber, Alex. 2001. "Idiolectal Error." *Mind and Language* 16: 263–83.
- Barwise, Jon, and John Etchemendy. 1987. *The Liar: An Essay on Truth and Circularity*. Oxford: Oxford University Press.
- Beall, Jc. 2001. "A Neglected Deflationist Approach to the Liar." *Analysis* 61: 126–29.
- , ed. 2008. *The Revenge of the Liar*. Oxford: Oxford University Press.
- . 2009. *Spandrels of Truth*. Oxford: Oxford University Press.

- Beall, Jc, and Julien Murzi. Forthcoming. "Two Flavors of Curry's Paradox." *Journal of Philosophy*.
- Beall, Jc, and Graham Priest. 2007. "Not So Deep Inconsistency." *Australasian Journal of Logic* 5: 74–84.
- Beall, Jc, and David Ripley. Forthcoming. "Non-Classical Truth Theories." In *The Oxford Handbook of Truth*, ed. M. Glanzberg. Oxford: Oxford University Press.
- Boghossian, Paul. 2006. "What Is Relativism?" In Greenough and Lynch 2006, 13–37.
- Brandom, Robert. 1994. *Making It Explicit*. Cambridge, MA: Harvard University Press.
- Burge, Tyler. 1979. "Semantical Paradox." *Journal of Philosophy* 76: 169–98.
- . 1982a. "Postscript to 'Semantical Paradox.'" In Martin, Robert L. 1984. *Recent Essays on Truth and the Liar Paradox*. London: Oxford University Press, 114–17.
- . 1982b. "The Liar Paradox: Tangles and Chains." *Philosophical Studies* 41: 353–66.
- Burgess, Alexis, and John Burgess. 2011. *Truth*. Princeton: Princeton University Press.
- Camp, Jr., Joseph L. 2002. *Confusion: A Study in the Theory of Knowledge*. Cambridge, MA: Harvard University Press.
- Cappelen, Herman. 2008. "Content Relativism and Semantic Blindness." In García-Carpintero and Kölbel 2008, 265–86.
- Cappelen, Herman, and John Hawthorne. 2009. *Relativism and Monadic Truth*. Oxford: Oxford University Press.
- . 2011a. "Reply to Lasersohn, MacFarlane, and Richard." *Philosophical Studies* 156: 449–66.
- . 2011b. "Reply to Glanzberg, Soames, and Weatherson." *Analysis* 71: 143–56.
- Chalmers, David. 2012. *Constructing the World*. Oxford: Oxford University Press.
- Chihara, Charles. 1979. "The Semantic Paradoxes: A Diagnostic Investigation." *Philosophical Review* 88: 590–618.
- . 1984. "The Semantic Paradoxes: Some Second Thoughts." *Philosophical Studies* 45: 223–9.
- Cobrerros, Pablo. 2011a. "Supervaluationism and Fara's Argument concerning Higher-Order Vagueness." In *Vagueness and Language Use*, ed. Paul Égré and Nathan Klinedinst, 207–21. New York: Palgrave Macmillan.
- . 2011b. "Varzi on Supervaluationism and Logical Consequence." *Mind* 120: 833–43.
- Cruse, D. A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Curry, H. 1942. "The Inconsistency of Certain Formal Logics." *Journal of Symbolic Logic* 7: 115–7.

- Davidson, Donald. 1984 [1973]. "Radical Interpretation." In *Inquiries into Truth and Interpretation*, 125–40. Oxford: Oxford University Press.
- . 1990. "The Structure and Content of Truth." *Journal of Philosophy* 87: 279–328.
- Devitt, Michael. 1996. *Coming to Our Senses*. Cambridge: Cambridge University Press.
- . 2006. *Ignorance of Language*. Oxford: Oxford University Press.
- Dummett, Michael. 1973. *Frege: Philosophy of Language*. Cambridge, MA: Harvard University Press.
- . 1991. *The Logical Basis of Metaphysics*. Cambridge, MA: Harvard University Press.
- Earman, John, and Arthur Fine. 1977. "Against Indeterminacy." *Journal of Philosophy* 74: 535–38.
- Egan, Andy. 2007. "Epistemic Modals, Relativism and Assertion." *Philosophical Studies* 133: 1–22.
- . 2010. "Disputing about Taste." In *Disagreement*, ed. Richard Feldman and Ted A. Warfield, 247–86. Oxford: Oxford University Press.
- Egan, Andy, and Brian Weatherson, eds. 2011. *Epistemic Modality*. Oxford: Oxford University Press.
- Egan, Andy, John Hawthorne, and Brian Weatherson. 2005. "Epistemic Modals in Context." In *Contextualism in Philosophy*, ed. Gerhard Preyer and Georg Peter, 131–70. Oxford: Oxford University Press.
- Einheuser, Iris. 2008. "Three Forms of Truth Relativism." In García-Carpintero and Kölbel 2008, 187–206.
- Eklund, Matti. 2002a. "Inconsistent Languages." *Philosophy and Phenomenological Research* 64: 251–75.
- . 2002b. "Deep Inconsistency." *Australasian Journal of Philosophy* 80: 321–31.
- . 2007. "Meaning-Constitutivity." *Inquiry* 50: 559–74.
- . 2008a. "The Liar Paradox, Expressibility, Possible Languages." In Beall 2008, 53–77.
- . 2008b. "Reply to Beall and Priest." *Australasian Journal of Logic* 6: 94–106.
- Fara, Delia Graff. 2011. "Truth in a Region." In *Vagueness and Language Use*, ed. Paul Égré and Nathan Klinedinst, 222–48. New York: Palgrave Macmillan.
- Feferman, Solomon. 1982. "Toward Useful Type-Free Theories, I." *Journal of Symbolic Logic* 49: 75–111.
- Field, Hartry. 1972. "Tarski's Theory of Truth." In Field 2001, 3–26.
- . 1973. "Theory Change and the Indeterminacy of Reference." In Field 2001, 177–93.
- . 1994. "Deflationist Views of Meaning and Content." In Field 2001, 104–40.
- . 2001. *Truth and the Absence of Fact*. Oxford: Oxford University Press.

- . 2006a. "Truth and the Unprovability of Consistency." *Mind* 115: 567–606.
- . 2006b. "Maudlin's *Truth and Paradox*." *Philosophy and Phenomenological Research* 73: 713–20.
- . 2008. *Saving Truth from Paradox*. Oxford: Oxford University Press.
- von Fintel, Kai, and Anthony Gillies. 2008. "CIA Leaks." *Philosophical Review* 117: 77–98.
- Friedman, Harvey, and Michael Sheard. 1987. "An Axiomatic Approach to Self-Referential Truth." *Annals of Pure and Applied Logic* 33: 1–21.
- Gaifman, Haim. 1992. "Pointers to Truth." *Journal of Philosophy* 89: 223–61.
- García-Carpintero, Manuel, and Max Kölbel, eds. 2008. *Relative Truth*. Oxford: Oxford University Press.
- Gillon, Brendan. 2004. "Ambiguity, Indeterminacy, Deixis, and Vagueness, Evidence and Theory." In Davis, Steven and Brendan Gillon. 2004. *Semantics: A Reader*. London: Oxford University Press, 157–90.
- Glanzberg, Michael. 2004. "A Contextual-Hierarchical Approach to Truth and the Liar Paradox." *Journal of Philosophical Logic* 33: 27–88.
- . 2007. "Context, Content, and Relativism." *Philosophical Studies* 136: 1–29.
- . 2011. "More on Operators and Tense." *Analysis* 71: 112–25.
- Greenough, Patrick, and Michael Lynch, eds. 2006. *Realism and Truth*. Oxford: Oxford University Press.
- Grover, Dorothy. 1977. "Inheritors and Paradox." *Journal of Philosophy* 74: 500–604.
- Gupta, Anil. 1982. "Truth and Paradox." *Journal of Philosophical Logic* 11: 1–60.
- . 1999. "Meaning and Misconceptions." In *Language, Logic, and Concepts*, ed. R. Jackendoff, P. Bloom, and K. Wynn, 15–42. Cambridge, MA: MIT Press.
- Gupta, Anil, and Nuel Belnap. 1993. *The Revision Theory of Truth*. Cambridge, MA: MIT Press.
- Halbach, Volker. 2011. *Axiomatic Theories of Truth*. Cambridge: Cambridge University Press.
- Hawthorne, John. 2007. "Craziness and Metasemantics." *Philosophical Review* 116: 427–40.
- Higginbotham, James. 1991. "Truth and Understanding." *Iyyun* 40: 271–88.
- Horn, L., and G. Ward. 2004. *The Handbook of Pragmatics*. Malden, MA: Blackwell.
- Horsten, Leon. 2011. *The Tarskian Turn*. Cambridge, MA: MIT Press.
- Jammer, Max. 2000. *Concepts of Mass in Contemporary Physics and Philosophy*. Princeton: Princeton University Press.
- Juhl, Cory F. 1997. "A Context-Sensitive Liar." *Analysis* 57: 202–4.
- Kaplan, David. 1989. "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals." In

- Themes from Kaplan*, ed. J. Almog, J. Perry, and H. K. Wettstein, 481–564. Oxford: Oxford University Press.
- Kennedy, Chris. 2010. "Ambiguity and Vagueness: An Overview." In *The Handbook of Semantics*, ed. C. Maienborn, K. Heusinger, and P. Portner, 507–34. Berlin: de Gruyter.
- Kölbel, Max. 2002. *Truth without Objectivity*. New York: Routledge.
- . 2003. "Faultless Disagreement." *Proceedings of the Aristotelian Society* 104: 53–73.
- Kremer, Philip, and Michael Kremer. 2003. "Some Supervaluation-Based Consequence Relations." *Journal of Philosophical Logic* 32: 225–44.
- Kripke, Saul. 1975. "Outline of a Theory of Truth." *Journal of Philosophy* 72: 690–716.
- Larson, Richard, and Gabriel Segal. 1995. *Knowledge of Meaning*. Cambridge, MA: MIT Press.
- Lasersohn, Peter. 2005. "Context Dependence, Disagreement, and Predicates of Personal Taste." *Linguistics and Philosophy* 28: 643–86.
- . 2008. "Quantification and Perspective in Relativist Semantics." *Philosophical Perspectives* 22: 305–37.
- . 2009. "Relative Truth, Speaker Commitment, and Control of Implicit Arguments." *Synthese* 166: 359–74.
- Lewis, David. 1983 [1979]. "Scorekeeping in a Language Game." In *Philosophical Papers*, vol. 1, 233–49. Oxford: Oxford University Press.
- . 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61: 343–77.
- MacFarlane, John. 2005a. "Making Sense of Relative Truth." *Proceedings of the Aristotelian Society* 105: 321–39.
- . 2005b. "The Assessment Sensitivity of Knowledge Attributions." *Oxford Studies in Epistemology* 1: 197–233.
- . 2007a. "Relativism and Disagreement." *Philosophical Studies* 132: 17–31.
- . 2007b. "The Logic of Confusion." *Philosophy and Phenomenological Research* 74: 700–8.
- . 2008. "Truth in the Garden of Forking Paths." In García-Carpintero and Kölbel 2008, 81–102.
- . 2009. "Nonindexical Contextualism." *Synthese* 166: 231–50.
- . 2011a. "Epistemic Modals Are Assessment-Sensitive." In Egan and Weatherson 2011, 144–78.
- . 2011b. "Simplicity Made Difficult." *Philosophical Studies* 156: 441–48.
- . Forthcoming. *Assessment Sensitivity: Relative Truth and Its Applications*.
- Matthews, Robert. 2007. *The Measure of Mind*. Oxford: Oxford University Press.
- Maudlin, Tim. 2004. *Truth and Paradox: Solving the Riddles*. Oxford: Oxford University Press.
- . 2006. "Replies." *Philosophy and Phenomenological Research* 73: 728–39.

- McGee, Vann. 1991. *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*. Indianapolis: Hackett Publishing Company.
- Montague, R. 1963. "Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability." *Acta Philosophica Fennica* 16: 153–67.
- Parsons, Charles. 1974. "The Liar Paradox." *Journal of Philosophical Logic* 3: 381–412.
- Patterson, Douglas. 2006. "Tarski, the Liar and Inconsistent Languages." *Monist* 89: 150–78.
- . 2007. "Inconsistency Theories: The Significance of Semantic Ascent." *Inquiry* 50, no. 6: 575–89.
- . 2009. "Inconsistency Theories of Semantic Paradox." *Philosophy and Phenomenological Research* 79: 387–422.
- Perry, John. 2001. *Reference and Reflexivity*. Stanford: CSLI.
- Petkov, Vesselin. 2009. *Relativity and the Nature of Spacetime*. Berlin: Springer.
- Pinillos, Ángel. 2010. "Time-Dilation, Context, and Relative Truth." *Philosophy and Phenomenological Research* 82: 65–92.
- Predelli, Stephano. 2005. *Contexts: Meaning, Truth, and the Use of Language*. Oxford: Oxford University Press.
- Priest, Graham. 1979. "The Logic of Paradox." *Journal of Philosophical Logic* 8: 219–41.
- . 2006a. *In Contradiction: A Study of the Transconsistent*, 2nd ed. Oxford: Oxford University Press.
- . 2006b. *Doubt Truth to Be a Liar*. Oxford: Oxford University Press.
- Priest, Graham, Jc Beall, and Bradley Armour-Garb, eds. 2005. *The Law of Non-Contradiction*. Oxford: Oxford University Press.
- Prior, Arthur. 1960. "The Runabout Inference Ticket." *Analysis* 21: 38–39.
- Recanati, François. 2007. *Perspectival Thought*. Oxford: Oxford University Press.
- Restall, Greg. 1994. "On Logics without Contraction." PhD thesis, The University of Queensland.
- . 2000. *An Introduction to Substructural Logics*. New York: Routledge.
- Richard, Mark. 2008. *When Truth Gives Out*. Oxford: Oxford University Press.
- Ripley, David. 2012. "Conservatively Extending Classical Logic with Transparent Truth." *Review of Symbolic Logic* 5: 354–78.
- . Forthcoming. "Paradoxes and Failures of Cut." *Australasian Journal of Philosophy*.
- Roberts, Craige. 2004. "Context in Dynamic Interpretation." In Horn and Ward 2004, 197–222.
- . 2010. "Retrievability and Definite Noun Phrases." Talk given at the philosophy department of The Ohio State University, February 5.
- . 2012 [1996]. "Information Structure: Toward an Integrated Formal Theory of Pragmatics." In *Semantics and Pragmatics* 5, article 6: 1–69.

- Schaffer, Jonathan. 2011. "Contextualism for Taste Claims and Epistemic Modals." In Egan and Weatherson 2011, 179–227.
- Scharp, Kevin. 2007. "Replacing Truth." *Inquiry* 50: 606–21.
- . 2008. "Aletheic Vengeance." In Beall 2008, 272–319.
- . 2011. "Xeno Semantics for Ascending and Descending Truth." In *Foundational Adventures: Essays in Honor of Harvey M. Friedman*, ed. Neil Tennant. Templeton Press (Online) and College Publications, foundationaladventures.com/foundational-adventures-essays-in-honor-of-harvey-m-friedman/kevin-scharp-xeno-semantics-for-ascending-and-descending-truth/.
- . 2013. *Replacing Truth*. Oxford: Oxford University Press.
- . Forthcoming. "Truth, Revenge, and Internalizability." *Erkenntnis*.
- Shapiro, Lionel. 2006. "The Rationale behind Revision-Rule Semantics." *Philosophical Studies* 129: 477–515.
- . 2011. "Expressibility and the Liar's Revenge." *Australasian Journal of Philosophy* 89: 297–314.
- Sider, Ted. 2009. "Ontological Realism." In *Metametaphysics*, ed. David Chalmers, David Manley, and Ryan Wasserman, 384–423. Oxford: Oxford University Press.
- Simmons, Keith. 1993. *Universality and the Liar: An Essay on Truth and the Diagonal Argument*. Cambridge: Cambridge University Press.
- Soames, Scott. 1999. *Understanding Truth*. Oxford: Oxford University Press.
- Sorensen, Roy. 2001. *Vagueness and Contradiction*. Oxford: Clarendon Press.
- Stalnaker, Robert. 1970. "Pragmatics." In Stalnaker 1999, 31–46.
- . 1974. "Pragmatic Presuppositions." In Stalnaker 1999, 47–62.
- . 1978. "Assertion." In Stalnaker 1999, 78–95.
- . 1998. "On the Representation of Context." In Stalnaker 1999, 96–114.
- . 1999. *Context and Content*. Oxford: Oxford University Press.
- Stephenson, Tamina. 2007. "Judge Dependence, Epistemic Modals, and Predicates of Personal Taste." *Linguistics and Philosophy* 30: 487–525.
- Tappenden, Jamie. 1993. "The Liar and Sorites Paradoxes: Toward a Unified Treatment." *Journal of Philosophy* 60: 551–77.
- Thomason, Richmond H. 1976. "Necessity, Quotation, and Truth: An Indexical Theory." In *Language in Focus*, ed. A. Kasher, 119–36. Dordrecht, Holland: D. Reidel.
- Ushenko, A. P. 1937. "A New 'Epimenides'." *Mind* 46: 549–50.
- van Fraassen, Bas C. 1968. "Presupposition, Implication, and Self-Reference." *Journal of Philosophy* 65: 136–52.
- Varzi, Achille. 2007. "Supervaluationism and Its Logics." *Mind* 116: 633–75.
- Weatherson, Brian. 2009. "Conditionals and Indexical Relativism." *Synthese* 166: 333–57.
- Williams, J. Robert G. 2007. "Eligibility and Inscrutability." *Philosophical Review* 116: 361–99.

- . 2008. "Supervaluationism and Logical Revisionism." *Journal of Philosophy* 105: 192–212.
- Williamson, Tim. 2000. "Semantic Paradox and Semantic Change." In *Proceedings of the Twentieth World Congress of Philosophy*, vol. 6, ed. A. Kanamori, 113–24. Bowling Green: Philosophy Documentation Center.
- Wright, Crispin. 2008. "Relativism about Truth Itself: Haphazard Thoughts about the Very Idea." In García-Carpintero and Kölbel 2008, 157–86.
- Yablo, Stephen. 1985. "Truth and Reflection." *Journal of Philosophical Logic* 14: 297–349.
- . 1993a. "Definitions, Consistent and Inconsistent." *Philosophical Studies* 72: 147–75.
- . 1993b. "Hop, Skip and Jump: The Agnostic Conception of Truth." *Philosophical Perspectives* 7: 371–96.
- . 1993c. "Paradox without Self-Reference." *Analysis* 53: 251–52.
- Zardini, Elia. 2011. "Truth without Contra(di)ction." *Review of Symbolic Logic* 4: 498–535.
- Zwicky, Arnold, and Jerrold Sadock. 1975. "Ambiguity Tests and How to Fail Them." In *Syntax and Semantics 4*, ed. J. Kimball Reimbald, 1–36. New York: Academic Press.

Derk Pereboom, *Consciousness and the Prospects of Physicalism*.
New York: Oxford University Press, 2011. x + 195 pp.

This splendid book makes several major contributions to the philosophy of mind. It is concerned principally with phenomenal consciousness, but it also has very useful things to say about a range of other topics, including representation, causation, causal exclusion, reduction, constitution, and the question of whether, and if so how, relational properties are grounded in intrinsic properties. It is innovative, deep, clear, judicious, and carefully argued, and it has many passages that illuminate classical texts and the contemporary literature. I will devote a section to each of its three main themes.

1. Introspective Awareness and Misrepresentation

In the first four chapters, Pereboom is principally concerned with a proposition he calls the *qualitative inaccuracy hypothesis*. According to this proposition, “our introspective representations fail to represent mental states as they are in themselves. More specifically, introspection represents phenomenal properties as having certain characteristic qualitative natures, and . . . these properties actually lack such features” (3). He does not assert that the qualitative inaccuracy hypothesis is true but rather maintains that it is an open epistemic possibility. As he shows, this is enough to put several key arguments for property dualism at risk, including the knowledge argument, the explanatory gap argument, and the Cartesian modal argument.

As can be seen from the foregoing formulation, the hypothesis rests on the doctrine that awareness of phenomenal properties is always representational in character: when one is aware of a phenomenal property, one’s state of awareness is partially constituted by a representation of the property. This doctrine implies that it is possible to draw a Kantian appearance/reality distinction with respect to awareness of phenomenal properties. On the one hand, there is the way that phenomenal properties are represented as being, and on the other, there is the way that phenomenal properties are in themselves.

I am grateful to Katherine Dunlop and James Van Cleve for help with the preparation of this review.

I will call the doctrine *Kantian phenomenal representationalism* (*KP-representationalism*, for short).

It is clear that KP-representationalism runs counter to many of the assumptions that are embedded in folk psychology. Thus, for example, folk psychology appears to deny that it is possible to draw an appearance/reality distinction with respect to awareness of pain. Moreover, the view is opposed by two of the main theories of awareness of phenomenal properties. On one of these theories, awareness of phenomenal properties involves an unanalyzable cognitive relation that Russell and others have called *acquaintance*. This relation is thought to be constitutively independent of representations and to provide immediate and full access to the essential natures of the items on which it is focused. On the other theory, which is generally known as *adverbialism*, it is a mistake to regard awareness of phenomenal properties as involving a relation of any sort. Rather, when we say, for example, that someone is aware of a pain in his or her foot, what we mean, or at least should mean, is that the person is aware of his or her foot painfully. In general, phenomenal properties are not objects of awareness but rather modes of being aware. Because KP-representationalism runs counter to so much of folk psychology and because it is opposed by these two well-known accounts of phenomenal awareness, it requires an elaborate defense. As far as I can tell, however, Pereboom does not argue for it in the present book. Presumably, for the present, at least, he wishes to maintain only that it is an epistemic possibility.

Pereboom's qualitative inaccuracy hypothesis can be seen as the conjunction of two doctrines. One is KP-representationalism and the other is a doctrine I will call the *misrepresentation thesis*. KP-representationalism allows for the possibility that our introspective representations are inaccurate. The misrepresentation thesis asserts that this possibility is in fact realized. Pereboom does not himself divide the qualitative inaccuracy thesis into these two components, but the distinction can be useful because the misrepresentation thesis can give rise to concerns that KP-representationalism does not occasion.

According to the misrepresentation thesis, introspective awareness represents phenomenal properties as having qualitative natures that they in fact lack. But what are these qualitative natures supposed to be? Are they natural properties? If so, then how can they create the illusion that phenomenal properties are nonnatural? And if not, then how can a physical mind acquire the ability to represent them? Further, as I understand him, whether qualitative natures are natural or not, Pereboom is committed to saying that when we use an introspective representation to attribute a qualitative nature to a phenomenal property, the attribution is always erroneous. But naturalistic theories of representation normally imply that representation involves a relationship of covariance between representations and the items to which they refer. It's not at all clear that this condition could be satisfied if introspective attributions of qualitative natures were as systematically erroneous as Pereboom appears to maintain.

It is possible that Pereboom is not obliged to rely as heavily on the misrepresentation thesis as he appears to do in practice. KP-representationalism provides a foundation for three quite different doctrines. One is the misrepresentation thesis itself. Another doctrine, which might be called the *underspecification thesis*, maintains that the representations in question underspecify the essential natures of phenomenal properties. Underspecification of this sort is a familiar feature of representation. For an example, consider a line drawing of a house. The drawing will inevitably leave out many of the details of the building's structure and provide no information at all about its material composition. Third, KP-representationalism provides a foundation for the *belief-fostering thesis*, according to which representations of phenomenal properties have a strong tendency to foster false beliefs about them, even though the representations themselves are not inaccurate. Thus, even though a representation refers to a complex property, it might be that the representation itself is syntactically simple and that it therefore functions within the larger cognitive system as a representational atom, affording no basis for higher cognitive faculties to appreciate the complexity of its referent. Reflection suggests that this doctrine and its immediate predecessor might be as useful in disarming arguments for dualism as the misrepresentation thesis.

Although the misrepresentation thesis figures much more prominently in the book than the two alternatives, Pereboom appears to have the underspecification thesis in mind in one or two passages, and he explicitly considers the belief-fostering thesis. More specifically, it figures in his treatment of the impression that phenomenal properties are primitive, in the sense that they are "metaphysically simple and thus not constituted by, in the sense of metaphysically analyzable into, a plurality of other properties" (16). Pereboom considers two explanations of the impression of phenomenal primitiveness. One is that the relevant representations misrepresent properties that are actually quite complex as being metaphysically simple. But another is that the representations foster a false belief to the effect that the represented items are primitive, perhaps because they are themselves syntactically simple. Pereboom indicates that he actually favors this second explanation.

Of course, it is one thing to explain why phenomenal states seem to us to be simple and quite another to explain why they seem to us to have qualitative natures that are altogether different from the natures of physical phenomena. As I understand him, Pereboom believes that it is necessary to rely on the misrepresentation thesis in giving an explanation of the latter sort. But it is not clear that this view is correct; it may be that the belief-fostering thesis will suffice. On one natural way of developing it, KP-representationalism implies that awareness of phenomenal states makes use of representations that are altogether different from the representations that figure in the awareness of physical states. Let us assume that this is the case. Is it any surprise that awareness that is supported by altogether different representations would give rise to an impression that the

represented phenomena are different? If the introspective system uses altogether different representations than the systems we use to represent physical states, it seems that it would be only natural for this difference to reverberate through our higher cognitive systems, resulting finally in an impression that introspectible phenomena are separated from physical states by an unbridgeable gulf.¹

2. Russellian Monism

As a first pass, *Russellian monism* can be characterized as making the following three claims: first, physics tells us nothing about the intrinsic characteristics of physical entities but is concerned only to describe their actual and potential relational properties; second, physical entities must nevertheless have intrinsic characteristics because it holds as a general rule that relations must be grounded in intrinsic natures; and third, the intrinsic characteristics of fundamental physical entities are capable of playing “a significant role in explaining consciousness or experience” (89). Pereboom devotes two chapters to articulating these core doctrines and to examining their epistemological credentials. His goal is to provide physicalists with a second strategy for defending their position in case the strategy outlined in section 1 does not work out.

Among other things, he considers various ideas about the nature of microphysical intrinsic characteristics. The two ideas that he takes most seriously are *phenomenal micropsychism* and *protophenomenalism*. Phenomenal micropsychism is a version of panpsychism—more specifically, it is the view that all fundamental physical entities, and therefore *all* physical entities, have intrinsic natures that are either the same as or very closely akin to the phenomenal properties that we encounter in experience. Protophenomenalism is the idea that the intrinsic properties of fundamental entities “are nonmental and similar enough to paradigmatic properties of current physics to count as physical, while at the same time they have a crucial role in grounding phenomenal properties of conscious states” (122).

After considerable discussion of these alternatives, Pereboom concludes that phenomenal micropsychism would require us to posit irreducible laws linking microphysical intrinsic properties to physical relations. These additional fundamental laws would add significantly to the complexity of Russellian monism, without providing any additional unity, and would therefore “fail to yield a deeply illuminating explanation of the [relational] properties specified by current microphysics” (115). On the other hand, there is some reason to hope that protophenomenalism can do better. To be sure, as Pereboom acknowl-

1. Like Pereboom, I am drawn to KP-representationalism, and I believe it has devastating consequences for dualism. I defend it in a recent book (Hill 2009) that complements Pereboom’s discussion.

edges, it is by no means clear at this point that it is possible for us to form adequate concepts of protophenomenal properties. Some philosophers, such as Colin McGinn, have argued that this is in fact the case: according to McGinn, we constitutionally lack the conceptual wherewithal to resolve the mind-body problem. But Pereboom joins Thomas Nagel and David Chalmers in being cautiously optimistic about the prospects of a satisfactory version of protophenomenalism. (As Pereboom emphasizes at length, Chalmers is largely responsible for bringing the protophenomenalist version of Russellian monism to the attention of contemporary philosophers. The term “protophenomenal properties” was devised by Chalmers.)

The case for Russellian monism requires us to agree that physics is not concerned with intrinsic properties, and this means that it requires us to accept that properties like *has a negative charge* are constituted by actual and potential relations. It is not clear to me that this is so. It is of course true that *has a negative charge* is intimately linked to relational properties, but I would have liked to see some concrete arguments that the relationship is constitutive rather than simply nomological. Moreover, the case for Russellian monism stands or falls with what Pereboom calls the Intrinsicness Principle:

Any mind-independently real substantial entity must have at least one substantial absolutely intrinsic property (101).

As Pereboom explains in detail, this principle has enjoyed considerable popularity in the history of philosophy, probably because there are widely shared intuitions to the effect that actual and potential relations must be *grounded in* intrinsic properties. It isn't clear that we should trust such intuitions, however, for it seems that there are clear counterexamples to a generalized grounding thesis. Thus, as Kant pointed out, things that are alike in their intrinsic properties can differ in point of their spatial relations. For example, drops of water that are duplicates of each other can be located at different positions in space and can stand at different distances from a third thing. Since the drops of water are intrinsic duplicates, it seems that neither of these spatial facts about the drops has anything to do with their intrinsic properties (Kant 1998, 368).

So far we have been reviewing general worries about the motivation for Russellian monism, but there is also a more particular concern about protophenomenalism. It seems to pose a dilemma. Either protophenomenal properties are sufficiently similar to familiar phenomenal properties that it is easy to see how they could constitute phenomenal properties, or they are not. On the one hand, if they are sufficiently similar, then it will be hard to see how they could provide a more-than-nomological ground for physical relations. Would it be at all explanatory to say, for example, that an intrinsic property that is similar to the phenomenal property *being a case of phenomenal yellow* provides a ground for the relational property *having a negative charge*? And on the other hand, if protophenomenal properties are quite different than phenomenal properties, then,

it seems, a philosopher who claims that phenomenal properties are constituted by protophenomenal properties is not in a much better epistemological or metaphysical position than a philosopher who claims that they are constituted by physical properties.

3. Nonreductive Physicalism

Pereboom devotes the last two chapters of the book to articulating and defending a novel version of nonreductive physicalism. On this view, mental state types are not identical with neural or functional state types but rather with abstract structural state types that can be exemplified both by concrete neural state tokens and by isomorphic state tokens in artificial brains. More specifically, mental state types are identical with *compositional* state types—abstract characteristics that “something has solely by virtue of intrinsic features of its parts and relations that those parts have to one another” (126–27). These compositional state types figure in causal laws and can therefore be said to bestow causal powers on their instances, but they differ from functional properties in that their association with causal powers is contingent. Pereboom’s theory also has an account of how tokens of mental states are related to tokens of neural states. The relation is constitution: tokens of mental states are constituted by tokens of neural states, which are in turn constituted by tokens of microphysical states. Thus, “in this position the physicality of mental state types is twice grounded, first by way of constitution of each token of the type in the microphysical, and second by way of identity to sufficiently abstract compositional properties” (127).

The resulting position is quite attractive, and Pereboom throws light on a number of related issues in developing it. I expect, however, that readers will have reservations about it.

In the first place, it is natural to have doubts about the motivation for nonreductive physicalism. For Pereboom as for others, the rationale for the view consists principally of the multiple realization argument. I suspect that many readers will have ceased to feel that this argument is rationally compelling.

To be sure, Pereboom does not endorse all versions of the argument; he keeps his distance from versions that are meant to show that members of other animal species have mental states like ours, despite having brains that are structurally quite different. I applaud this skepticism; those versions of the argument are in my view extremely weak. (For one thing, the intuitive attributions of mental states to which they appeal can be explained away as the products of a quick and very dirty System One heuristic that is activated principally by biological motion. These intuitions are withdrawn when a reflective, System Two process is given information about the usually rather substantial divergences between the human brain and the brains of the relevant animals.²) As against

2. For an extended discussion of the differences between System One and System

Pereboom, however, I think there are also grounds for questioning versions of the multiple realization argument that are concerned with androids whose brains are composed of silicon but are nonetheless structurally isomorphic with ours. It is an immediate consequence of Block's thought experiment involving the Chinese nation that our attributive practices do not recognize structural isomorphism as a sufficient ground for attributing phenomenal properties. (In Block 1978, the Chinese nation is organized in a way that reflects the abstract neural structure of the human brain as well as its functional properties. Most people who consider this example are unwilling to say that the Chinese nation enjoys phenomenal consciousness.) Pereboom might reply by urging that we have an inclination to attribute phenomenal states to embodied androids like C3PO, at least when we have reason to think that their brains are structurally similar to ours. There is some merit in this claim, but I think that the inclination has more to do with C3PO's capacity for speech and biological motion than the imagined isomorphism; for if it derived primarily from the isomorphism, we would be willing to attribute conscious mental states to billions of Chinese people in a stadium, given that the interactions of the individual people are akin to the interactions of human neurons.

In general, I would have a lot more confidence in Pereboom's metaphysical system-building if it were grounded in some x-phi data concerning intuitions about androids under various scenarios. For example, are people less inclined to attribute conscious experiences to C3PO after they have considered Block's example? What happens to their intuitions when C3PO's powers of speech and biological motion are reduced because his brain is housed in a different sort of body? What happens to them in contexts in which the differences between silicon chips and human neurons have been accentuated? My guess is that when answers to these questions are in hand, they will not be such as to mandate nonreductive versions of physicalism, though they may leave it as an option.

A second concern is that all versions of the central state identity thesis, including ones that affirm the identity of mental state types with abstract structural properties of neural events, are challenged by transparency considerations. Qualia seem to be out there in the body, or to be properties of external objects, albeit properties that can be experienced only by viewers with perceptual systems of a certain sort. The pain is in my foot, and the way that banana looks right now is a relational property of the banana. All versions of the central state identity thesis, including Pereboom's theory, must reject these fundamental intuitions. In saying this, I don't mean to deny that central state

Two processes, see Kahneman 2011. For a demonstration that biological motion is sufficient to trigger a quick and dirty heuristic for attributing mental states to other beings, see the classic video by Heider and Simmel 1944.

identity theories have a lot going for them. On the contrary, I feel that they receive strong support from the fact that there are very tight correlations linking conscious experiences with brain states. My view is that all theories of qualia are faced with a paradox, in which intuitions of transparency are pitted against correlational data. Before I could embrace Pereboom's theory, I would have to see something that he has not yet provided—a resolution of the paradox that explains away transparency intuitions and sustains the view that correlations are best explained by identity.³

4. Conclusion

Although Pereboom's book leaves us with many questions, it is an outstanding work that philosophers of mind of all stripes will find richly rewarding. I have ordered it for my fall seminar and will no doubt use it again in future courses.

References

- Block, N. 1978. "Troubles with Functionalism." In *Perception and Cognition: Issues in the Foundations of Psychology*, ed. C. W. Savage, 261–325. Minnesota Studies in the Philosophy of Science, vol. 9. Minneapolis: University of Minnesota Press.
- Heider, F., and M. A. Simmel. 1944. "Heider–Simmel Demonstration." [www. youtube. com/](http://www.youtube.com/).
- Hill, C. S. 2009. *Consciousness*. Cambridge: Cambridge University Press.
- . 2012. "Locating Qualia: Do They Reside in the Mind or in the Body and the World?" In *New Perspectives on Type-Identity*, ed. S. Gozzano and C. S. Hill, 127–49. Cambridge: Cambridge University Press.
- Kahneman, D. 2011. *Thinking Fast and Slow*. New York: Farrar, Strauss, Giroux.
- Kant, I. 1998. *Critique of Pure Reason*, ed. P. Guyer and A. W. Wood. Cambridge: Cambridge University Press.

Christopher S. Hill
Brown University

Philosophical Review, Vol. 122, No. 3, 2013
DOI 10.1215/00318108-2087867

3. I discuss this paradox in Hill 2012. According to the resolution I endorse there, while phenomenal awareness is tightly correlated with brain states, the objects of such awareness are extracranial.

Casey Perin, *The Demands of Reason: An Essay on Pyrrhonian Scepticism*. Oxford: Oxford University Press, 2010. 130 pp.

To give a complete and completely consistent account of Pyrrhonian skepticism as it is found in the pages of Sextus Empiricus is no small task, perhaps even a Sisyphean one. It is for just this reason that, in *The Demands of Reason*, Casey Perin distances himself from any such endeavor, freely admitting that his account of Pyrrhonian skepticism is not a comprehensive one (5–6). Instead, what Perin sets out to do, and does well, is to focus on four aspects of Pyrrhonian skepticism that emerge from a consideration of Sextus's *Outlines of Pyrrhonism* (hereafter *PH*) that he thinks are of particular philosophical significance.¹ They are: the issue of whether the skeptic is committed to searching for truth; the issue of whether the skeptic is committed to certain principles of rationality; the scope of the skeptic's skepticism; and the connection between skepticism and action. With respect to these four areas, Perin argues for the following four theses, devoting a chapter to the articulation and defense of each:

- (1) The skeptic is committed to the search for truth.
- (2) The skeptic is committed to certain principles of rationality.
- (3) The skeptic has beliefs about the way things appear to him or her to be but lacks beliefs about the way things are.
- (4) The skeptic is able to act.

Perin writes clearly, concisely, and often persuasively in defense of (1)–(4). While I cannot here do justice to all his thought-provoking claims, I shall offer some comments on the general argumentative structure of the book, followed by two critical reflections on (1) and (2), the book's most interesting and controversial claims.

There is a natural divide in *The Demands of Reason*. (1) and (2) go together as do (3) and (4). More specifically, for Perin, (2) is true in virtue of (1)'s being true, and (4) is true in virtue of (3)'s being true. Consequently, for his account to hang together, not only does Perin have to marshal reasons for thinking his pivotal claims (1) and (3) are true, he also has to lay out exactly how the truth of (2) can be derived from the truth of (1) and how the truth of (4) can be derived from the truth of (3). Perin's attempt to meet these desiderata constitutes the bulk of *The Demands of Reason*.

Perin acquits himself well when it comes to arguing for the truth of (3). He focuses on *PH* 1.13, a crucial stretch of argument that has been the crux of many a Sextan interpretation, and offers a sustained critique of Michael Frede's

Thanks to Gail Fine and David Charles for their helpful comments.

1. Sextus Empiricus, *Opera*, ed. H Mutschmann and J. Mau, 4 vols. (Leipzig: Teubner, 1954–56).

(1998) influential reading of the passage. Roughly, whereas Frede distinguished between those beliefs a skeptic could have and those beliefs he had to lack on the basis of the grounds upon which the beliefs were held, Perin draws the distinction on the basis of the content of the beliefs: those beliefs a skeptic can have are beliefs about how things appear to the skeptic to be, whereas those beliefs a skeptic has to lack are those beliefs about how things are (62). It should be noted that Perin's path has been trodden before—the explorations of Gail Fine (2000) spring to mind—but Perin is aware of this fact and draws attention to it during his careful critique of Frede's position. Perin then goes on to analyze how the denial of (4) stands or falls depending on what the denier of (4) means by "belief." Perin argues that the nondoxastic appearances that the skeptic has (and indeed cannot avoid having), when coupled with desire, are sufficient for action. Hence, the skeptic is able to act even if he or she lacks beliefs (96–97).

(1) and (2) are more original and controversial theses. They provide the central theme of the book, which is that the Pyrrhonian skeptic is not an antirationalist,² but he or she aims to satisfy the eponymous demands of reason. Perin offers a variety of arguments in favor of (1). His opening moves are cautious. He begins by analyzing a number of Sextan texts to show that they need not be taken as evidence for the non-truth-directed nature of the skeptic's inquiry (10–20). He then becomes bolder and offers a positive account of how engaging in truth-directed inquiry is compatible both with the skeptic's goal of achieving tranquillity through the suspension of judgment (20–27) and the skeptic's use of the Agrippan modes (27–32).

Following this, Perin offers two arguments for (2). His first is linguistic. Beginning with the datum that at *PH* 2.96 and *PH* 3.36 Sextus uses the locution "it is right to suspend judgment" interchangeably with the locution "it is necessary to suspend judgment," Perin argues that, even if there are a number of passages in the *Outlines* that are neutral between interpretations that take the necessity to be psychological and those that take it to be grounded in the requirements of rationality (*PH* 1.61, 123, 128, 140, 170), in at least *PH* 2.96 and 3.36, the necessity is to be understood not merely in the former but also in the latter sense (47–49). Perin's second argument is a priori and depends on the truth of (1). It rests on the plausible claim that an epistemic subject cannot be engaged in the search for truth without aiming to satisfy certain basic rational requirements, for instance, the requirement that one should suspend judgment about whether P if one believes there is no reason to believe either P or its negation (52–56).

That gives a sense of the component parts of Perin's interpretation. I end by offering two brief critical remarks regarding (1) and (2). First, in his articulation and defense of (1), Perin claims that an essential feature of skeptical

2. Perin ascribes this view to commentators like John Palmer and Gisela Striker. See, by way of comparison, Palmer (2000) and Striker (2001).

inquiry is that it is directed at discovering whether some proposition or other is true (8). It would, however, have been illuminating were Perin to have elucidated what *kind* of proposition this is. A piece of wax melts in the heat of the sun.³ Does the skeptic investigate whether the heat of the sun has caused that piece of wax to melt? Does he or she investigate whether he or she is justified in believing that the heat of the sun has caused that piece of wax to melt? Does he or she investigate whether the concept of causation is a coherent one? Perin writes as if the skeptic is concerned only with the first kind of truth-directed investigation. But there is evidence from the *Outlines* that suggests that Sextus's skeptic also had an eye on the other two kinds, namely, on investigation that consists in determining whether a given epistemic subject is justified in believing what he or she believes (*PH* 1.20, 177, 186, 205, 212, 237), and investigation that consists in determining whether a given dogmatic concept, for example, causation, is coherent (*PH* 3.13–16).⁴ It would have been welcome if Perin had distinguished between these different kinds of truth-directed investigation and had himself investigated which (if any) was the most basic form of inquiry at play in Sextus's writings.

Second, Perin's textually based argument in favor of (2) needs further support. The Greek impersonal verb *prosēkei* has as its root meaning "it belongs to" or "it concerns," not "it is right," so to argue that the term has the kind of normative force that Perin takes it to have requires a more careful examination of the uses of the expression elsewhere in the *Outlines* (*PH* 1.86, 2.42, 54, 60, 77, 3.29, 78, 82, 150, 234, 257) and in philosophical Greek more generally. For instance, just restricting oneself to the *Outlines*, some account should have been given of the use of the locution at *PH* 1.87, where a natural way to understand it is in terms of moral obligation, and likewise at *PH* 3.78, 82, and 150, where a natural way is in terms of logical necessity.

With all that being said, Perin's anti-anti-rationalist interpretation of Sextus has much to commend it. *The Demands of Reason* demands reading by anyone interested in Pyrrhonian skepticism.

References

- Fine, G. 2000. "Sceptical Dogmata: Outlines of Pyrrhonism 1 13." *Metaphis* 13: 81–105.
 Frede, M. 1998. "The Sceptic's Two Kinds of Assent and the Question of the Possibility of Knowledge." In *The Original Sceptics: A Controversy*, ed. M. F. Burnyeat and M. Frede, 127–51. Indianapolis: Hackett.

3. The example is Sextus's. See *PH* 3.14.

4. See, by way of comparison, *PH* 2.22–28.

Palmer, J. 2000. "Skeptical Investigation." *Ancient Philosophy* 20: 351–73.

Striker, G. 2001. "Scepticism as a Kind of Philosophy." *Archiv für Geschichte der Philosophie* 83: 113–29.

Stefan Sienkiewicz

Oriel College, Oxford

Philosophical Review, Vol. 122, No. 3, 2013

DOI 10.1215/00318108-2087876

Gordon Belot, *Geometric Possibility*.

Oxford: Oxford University Press, 2011. x + 219 pp.

This is a neat little book (138 pages without appendices, approximately 200 with). It focuses on one aspect of the debate between substantivalism and relationalism about space. (Belot discusses things in terms of space rather than spacetime, explaining this choice at the end.) As Belot convincingly argues in chapter 2, the relationalist should be a modal relationalist, positing, in the spatial facts about a world, not just the actual configurations of material objects but the possible ones too. Others have argued that the relationalist should go modal, without tackling the difficult question of exactly what kind of modality is involved. *Geometric Possibility* is a sustained look at what this kind of modality could be. Belot explores three different accounts of this notion, which fall in line with three different accounts of laws. The result is a book that has interesting things to say not only about modal relationalism and different spatial geometries but also about laws of nature, varieties of modality, and ontology generally.

The introduction is primarily stage-setting, but it also makes an important point: the relationalist, like the substantivalist, can be a realist in the sense of "attribut[ing] to reality a determinate spatial structure" (1). For the relationalist, this is the structure of the (possible and actual) spatial relations among material objects; for the substantivalist, this is the structure of relations among the parts of space.

This yields a somewhat different understanding of the debate from what's typically given: "Relationalists and substantivalists can agree... that space is a thing of some sort and that it has some geometric structure. Their disagreement concerns the nature of the existence of space" (2). I largely agree. However, Belot doesn't want to frame the dispute as about ontology, whether space *exists*, since he thinks this blurs the realism/antirealism distinction, and here we part company. I'd say this instead. Both the relationalist and substanti-

valist can be realists about spatial structure, as Belot notes. Still, they do disagree about whether space exists: their dispute is about what ontology underlies that spatial structure. The substantivalist says this structure inheres in a physical space. The relationalist denies that there exists any such physical space, saying this is simply a structure of relations among material objects. To me, this best captures the characteristic difference between the two views, while acknowledging that the relationalist can be a kind of spatial realist.

In chapter 1, Belot explores the question of which mathematical structures correspond to the spatial geometries of possible worlds. He begins with the reasonable premise that Euclidean space is one possible structure and suggests that generalizations of this correspond to other possibilities. He concludes, "All possible spatial geometries are represented by metric spaces" (31), spaces with a natural notion of distance, and he describes a variety of these. Many of the examples pop up later in the book, though some readers may not be as interested in the mathematical details until after reading later chapters. Particularly interesting here is Belot's objection to the often tacit assumption that Riemannian manifolds give us all the possible spatial structures. Limiting possible spaces to these structures violates a natural principle of plenitude since the space of Riemannian manifolds is "gappy."

Chapter 2 contains Belot's arguments that the relationalist should be a modal relationalist. The overarching idea is that there are even quite simple spatial structures we surely wish to countenance, yet which the nonmodal relationalist is unable to capture. Modal relationalism provides the truth conditions needed for the nonsubstantivalist to make sense of claims about the spatial geometry of a world.

At the end of the chapter, Belot shows that three initially plausible conditions on geometric possibility cannot (given one further natural assumption) be jointly satisfied. In following chapters, he evaluates three accounts of geometric possibility, each rejecting one of these conditions. Chapter 3 discusses Huggett's (2006) relationalism, which goes naturally with a best-system account of both laws and geometric possibility. Chapter 4 discusses primitivism about geometric possibility, analogous to primitivism about laws. Chapter 5 discusses a view of geometric possibility analogous to necessitarianism about laws.

The conclusion sums up and discusses the (often obscure) relationship between relationalism about ontology and about motion. Belot concludes that either primitivism or necessitarianism is most plausible, without taking a firm stand on either one, although he most likes the latter. Either way, the relationalist requires a new notion of geometric possibility, which is distinct from physical possibility and may not, surprisingly, be grounded in the ordinary geometric properties instantiated at a world (such as the distance relations among material objects).

Finally, there are five appendices. In one, Belot argues that the simplicity of an ontology (though a relationalist, he seems to admit that substantivalism is simpler in this way) isn't evidence that it is true.

I'll end with three general thoughts differing from Belot's. (There is much more worth discussing.) First, it seems to me that there's a more straightforward argument that modal relationalism is the most viable form of relationalism. The relationalist (as well as the substantivalist) should posit the spacetime structure needed for the laws—in accord with a generally accepted methodological principle—and the actually instantiated spatiotemporal relations won't, in general, suffice to fix or ground that structure.

Second, I wonder whether modal relationalism requires a *sui generis* kind of modality. In chapter 2, Belot gives two arguments against the view that this is just physical possibility. First, there are generally different global spatial structures allowed by a set of laws; for example, both infinite Euclidean space and a three-torus are consistent with Maxwell's equations and the Lorentz force law. So at any world allowed by these laws, it's physically possible for there to be an infinite array of material particles—that is, for some particles to be infinitely far apart—even though this isn't geometrically possible at the toroidal worlds. Second, the relationalist needs a distinct kind of modality to accommodate claims about certain spatial structures, for example, that space can be infinite even if it's physically impossible for matter to travel beyond a finite region.

Though reasonable, these arguments aren't, I think, decisive. Against the second, the relationalist (and even the substantivalist) could deny that such a scenario makes sense. If material objects cannot move beyond a certain region, perhaps that region is all there is to the spatial structure of the world—especially if the laws say it's impossible for particles to travel to infinity, and one thinks the laws indicate a world's spatial structure.

Against the first, suppose that physical possibility is consistency with the laws (the details depend on one's view of laws). Generally, a set of laws will require a certain spacetime structure. For example, the laws of special relativity require a Minkowski spacetime: other spacetime metrics aren't physically possible. Even so, there may be other differences in (spatial or spacetime) structure allowed by the laws, such as differences in global spatial topology. Now consider all the worlds consistent with some laws. Relative to any one of these worlds, the others are physically possible in that they're allowed by the world's laws, even though some won't be (physically) possible once we add further physical facts about the given world, such as the global topology, as constraints. Thus, relative to Belot's toroidal world, there's a sense in which an infinite array of particles is physically possible: there are other worlds consistent with the laws that have such arrays. But there's also a sense in which this isn't physically possible since within the subset of possible worlds whose space is toroidal, there can't—according to the spatial geometry and the laws that allow that structure—be infinite arrays of particles. Perhaps there is simply a broader and narrower sense of physical

possibility in play here, depending on how much one specifies about a world, rather than a distinct kind of possibility. Compare: specifying a world's initial conditions further constrains the behaviors seen as physically possible relative to the world, yet we don't infer a distinct kind of "initial-conditions possibility." Maybe what Belot calls geometric possibility isn't a new, *sui generis* modality but a kind of physical possibility—what's possible according to the laws, which themselves constrain the geometry of space(time). (Perhaps in the end there is no way for the relationalist to make this work, in which case this is more of an argument for substantivalism than an objection to the idea that the relationalist needs a new kind of possibility.)

Third, Belot objects to a Quinean view that in ontological investigations, we should think the simplest theory likely to be true, though he allows for this in scientific inquiry. If one takes as a starting point that ontological inquiry is continuous with scientific inquiry, however, one will thereby think their methodologies alike, and Belot's arguments to the contrary will be nonstarters. Belot further argues that in ontological inquiry, simplicity isn't evidence of truth because it isn't sensitive to truth: we'd have the same evidence, including the same philosophical arguments, in a world with a given spatial structure regardless of whether space is substantival or relational there. However, I'm not sure that we need more than the actual track record of the methodological principles we use (including parsimony principles), which we think have generally been successful—even if we can't show that they reliably track truth. Indeed, even if relationalism and substantivalism are each compatible with the evidence that can be obtained by a world's inhabitants, I think this just shows that we can be wrong in our inference to a given ontology, not that we can't still have good reason to infer the simpler ontology.

Altogether, this is a very good book, which anyone interested in space-time ontology and related issues in philosophy of physics, philosophy of science, and metaphysics should read.

Reference

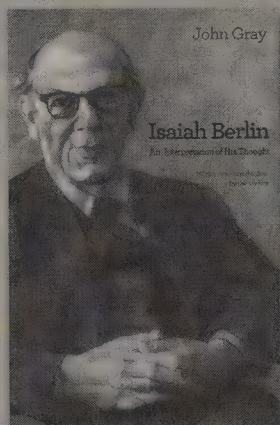
Huggett, Nick. 2006. "The Regularity Account of Relational Spacetime." *Mind* 115: 41–73.

Jill North

Cornell University

Philosophical Review, Vol. 122, No. 3, 2013

DOI 10.1215/00318108-2087885



Isaiah Berlin

An Interpretation of His Thought

John Gray

With a new introduction by the author

"An acute and illuminating exposition of Berlin's world view. . . . [Gray] probably gets closer to Berlin than anyone else has done."

—Michael Walzer, *New York Review of Books*

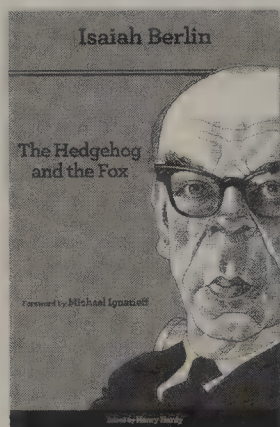
"A masterly study of Berlin's political thought."

—Adam Wolfson, *National Interest*

"An excellent exposition of Isaiah Berlin's ideas."

—William Phillips, *Partisan Review*

Paper \$22.95 978-0-691-15742-9



The Hedgehog and the Fox

An Essay on Tolstoy's View of History

Second Edition

Isaiah Berlin

Edited by Henry Hardy

With a new foreword by Michael Ignatieff

"Beautifully written and suggestive."

—W. H. Auden, *New Yorker*

"A brilliant essay . . . a searching and profound analysis."

—E. H. Carr, *Times Literary Supplement*

"So entertaining, as well as acute, that the reader hardly notices that it is learned too."

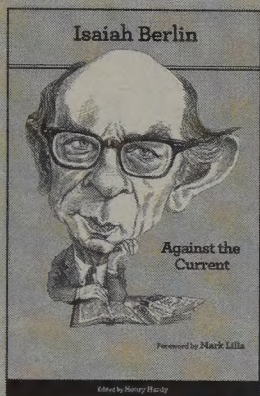
—Arnold Toynbee, *Observer*

Paper \$12.95 978-0-691-15600-2



PRINCETON UNIVERSITY PRESS

See our E-Books at
press.princeton.edu



Against the Current

Essays in the History of Ideas

Second Edition

Isaiah Berlin

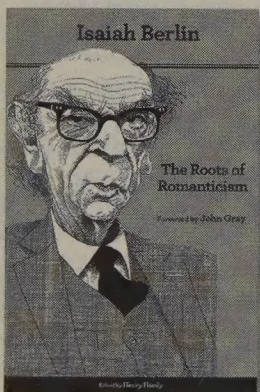
Edited by Henry Hardy

With a new foreword by Mark Lilla

"Berlin expounds the ideas of half-forgotten thinkers with luminous clarity and imaginative empathy . . . exhilarating to read."

—Keith Thomas, *Observer*

Paper \$24.95 978-0-691-15610-1



The Roots of Romanticism

Second Edition

Isaiah Berlin

Edited by Henry Hardy

With a new foreword by John Gray

"A fascinating intellectual history."

—Douglas A. Sylva, *New York Times Book Review*

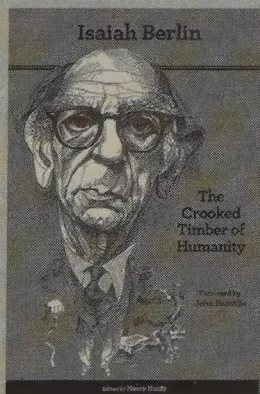
A. W. Mellon Lectures in the Fine Arts

National Gallery of Art, Washington

Bollingen Series XXXV: 45

Paper \$12.95 978-0-691-15620-0

Published in association with the National Gallery of Art, Washington



The Crooked Timber of Humanity

Chapters in the History of Ideas

Second Edition

Isaiah Berlin

Edited by Henry Hardy

With a new foreword by John Banville

"A beautifully patterned tapestry of philosophical thought. . . A history of ideas that possesses all the drama of a novel, all the immediacy of headline news."

—*New York Times*

Paper \$24.95 978-0-691-15593-7

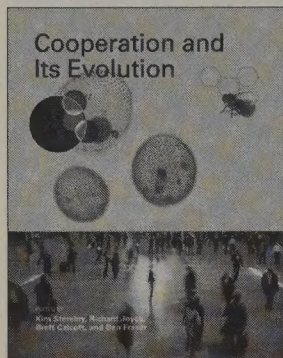


PRINCETON UNIVERSITY PRESS

See our E-Books at
press.princeton.edu



The MIT Press /// Philosophy



COOPERATION AND ITS EVOLUTION

edited by Kim Sterelny, Richard Joyce, Brett Calcott, and Ben Fraser

Essays from a range of disciplinary perspectives show the central role that cooperation plays in structuring our world.

Life and Mind series • A Bradford Book
608 pp., 11 illus., \$55 cloth

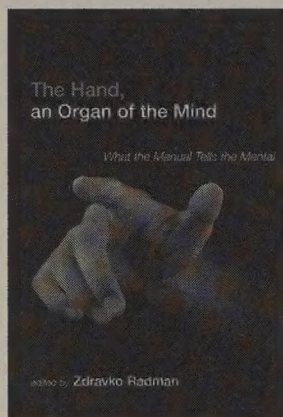
COMMUNICATING MORAL CONCERN

An Ethics of Critical Responsiveness

Elise Springer

A novel reframing of moral agency, emphasizing the responsive habits and skills by which we engage one another's attention to moral concerns.

328 pp., 1 illus., \$39 cloth



THE HAND, AN ORGAN OF THE MIND

What the Manual Tells the Mental

edited by Zdravko Radman

Theoretical and empirical accounts of the interconnectedness between the manual and the mental suggest that the hand can be understood as a cognitive instrument.

464 pp., 8 illus., \$50 cloth

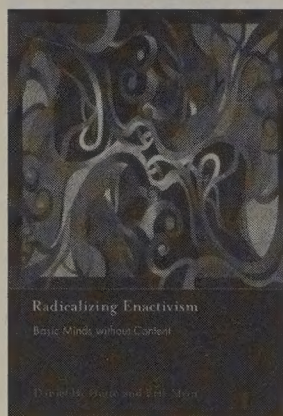
COMPUTABILITY

Turing, Gödel, Church, and Beyond

edited by B. Jack Copeland, Carl J. Posy, and Oron Shagrir

Computer scientists, mathematicians, and philosophers discuss the conceptual foundations of the notion of computability as well as recent theoretical developments.

376 pp., 1 illus., \$35 cloth



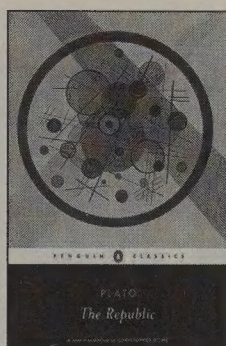
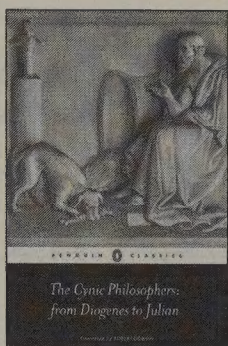
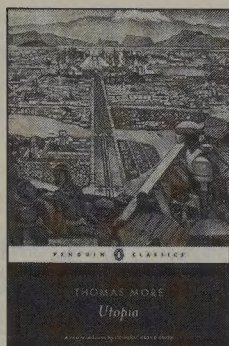
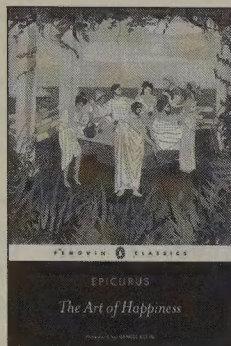
RADICALIZING ENACTIVISM

Basic Minds without Content

Daniel D. Hutto and Erik Myin

A book that promotes the thesis that basic forms of mentality—intentionally directed cognition and perceptual experience—are best understood as embodied yet contentless.

248 pp., 1 illus., \$35 cloth



EPICURUS THE ART OF HAPPINESS

*Translated with an Introduction
and Commentaries by George K. Strodach
Foreword by Daniel Klein*

This volume places Epicurus's extant writings—his letters, doctrines, and Vatican sayings—alongside parallel passages from Lucretius, the greatest exponent of his philosophy, and includes a lucid introductory essay about his teachings.

Penguin Classics • 272 pp. • 978-0-14-310721-7 • \$16.00

DANIEL KLEIN TRAVELS WITH EPICURUS A Journey to a Greek Island in Search of a Fulfilled Life

One of the bestselling authors of *Plato and a Platypus Walk into a Bar* travels to Greece, where he learns how to age well by returning to the wisdom of renowned philosophers, from Epicurus in antiquity to Sartre in modern times. "Witty and wry."—*The Huffington Post*.

Penguin • 176 pp. • 978-0-14-312193-0 • \$20.00

THOMAS MORE UTOPIA

*Newly Translated with an Introduction
by Dominic Baker-Smith*

A major new translation of More's famous work of philosophical fiction that satirizes Europe of the 16th century by imagining an alternative society enjoying freedom from fear, oppression, violence, and suffering.

Penguin Classics • 192 pp. • 978-0-14-144232-7 • \$12.00

ROBERT DOBBIN, editor & translator THE CYNIC PHILOSOPHERS From Diogenes to Julian

This unique new anthology illuminates the ideals of the ancient Greek and Roman Cynics by drawing together the writings on and by various Cynic philosophers, from founding figures Antisthenes and Diogenes to Hipparchia, one of the few female philosophers in antiquity, to 4th-century Roman emperor Julian the Apostate.

Penguin Classics • 352 pp. • 978-0-14-119222-2 • \$17.00

CICERO ON LIVING AND DYING WELL

*Newly Translated with an Introduction
and Notes by Thomas Habinek*

The philosophical treatises on such diverse and trenchant topics as friendship, religion, death, fate, and scientific inquiry by the great Roman orator are now available in a vital new translation that renders his writings accessible to modern readers as never before.

Penguin Classics • 240 pp. • 978-0-14-045556-4 • \$16.00

PLATO THE REPUBLIC

*Newly Translated with an Introduction
and Notes by Christopher Rowe*

An authoritative new translation of Plato's foundational work of Western philosophy that creates a picture of an ideal society governed not by money, power, or fame, but by philosophy, wisdom, and justice.

Penguin Classics • 496 pp. • 978-0-14-144243-3 • \$13.00



PENGUIN GROUP (USA)
Academic Marketing Department

375 Hudson Street New York, New York 10014

www.penguin.com/academic

IN THIS ISSUE

Epistemic Teleology and the Separateness of Propositions

Selim Berker

Peer Disagreement, Evidence, and Well-Foundedness

Han van Wietmarschen

Truth, the Liar, and Relativism

Kevin Scharp

BOOK REVIEWS